1971-09

A comparison of short-term forecasting models.

Hayes, Ralph Eugene.

Monterey, California ; Naval Postgraduate School

http://hdl.handle.net/10945/15611

A COMPARISON OF SHORT-TERM FORECASTING MODELS

Ralph Eugene Hayes

United States
Naval Postgraduate School

THESIS

A COMPARISON OF SHORT-TERM
FORECASTING MODELS

by

Ralph Eugene Hayes

Thesis Advisor: P. W. Zehna

September 1971

A Comparison of Short-Term

Forecasting Models

by

Ralph Eugene Hayes
Major, United States Army
B.S., North Georgia College, 1960

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1971

ABSTRACT

Seven short-term forecasting models, two using least squares estimation methods and five employing variations of the exponentially weighted moving average method, are compared in their relative ability to produce minimum error variance forecasts for seven simulated time series. Each series was generated to enable one of the forecast models to be the least squared error predictor. A comparison methodology is developed which facilitates forecast model selection based on single or group series forecast performance through the measurement of model specification errors. A computer program is presented which may be modified to accept real time series and which permits the forecast models to be ranked in order of their relative specification error.

TABLE OF CONTENTS

TABLE OF SYMBOLS AND ABBREVIATIONS

Following are the meanings of the notation most frequently used throughout this thesis.

$\hat{Y}_t$ ≡ the estimate or forecast of the observed value (Y_t) of a random process $\{Y_t\}$ at time t.

$\hat{\bar{Y}}_t$ ≡ the estimate or forecast of the level of mean value ($\bar{Y}_t$) at time t of the process $\{Y_t\}$ made at time t.

$\hat{b}_t$ ≡ the estimate at time t-1 of the process slope (b_t) at time t in a linear model.

ε_t ≡ a random shock or superimposed error experienced by the random process at time t which contributes to the error in forecasting the observed value.

F_t ≡ the one step ahead forecast at time t-1 of the as yet unobserved series value, Y_t.

e_t ≡ the one step ahead forecast error ($e_t = Y_t - F_t$) which is realized when the observation Y_t is taken.

α ≡ the smoothing constant in a EWMA model. Also, α is the portion of a random shock which is considered to be a permanent contribution to process level. $\alpha = 1-\beta$.

β ≡ the discounting factor in an EWMA model; the rate at which the weight given to past observations diminishes. $\beta = 1-\alpha$.

α' ≡ the smoothing constant associated with the slope estimator of a forecast model.

β' ≡ the discounting factor associated with the forecast model slope estimator. β' is not necessarily equal to $1-\alpha'$.

$\hat{\beta}_2$ $\equiv$ the estimate of β_2, the slope of the linear model in least squares methodology.

ρ $\equiv$ the autocorrelation coefficient of the series $\{Y_t\}$ with lag of one period.

f $\equiv$ the measure of model specification error.

I. INTRODUCTION

When a random event has a significant impact on society, considerable effort will be made to predict its occurrence. The ability to predict or forecast such an event is a by-product of a quantitative understanding of the situation, a physical model. Prediction based upon a behavioral model is the ideal, but forecasts can also be based upon recognition of regularity as well as upon explanation of that regularity.

There are many situations in both industrial and governmental operations which require forecasts for hundreds or thousands of routinely recurring events. Often these events, such as equipment failures which generate demands for repair parts, are not influenced by advertising or other factors. In the absence of additional information a projection into the future must be made based entirely upon past observations of these events. It is frequently true that none of the events, individually, are of sufficient importance to warrant the study and attention required to develop behavioral models. For these items of low cost and nonsensitive nature a routine forecasting system is desired which employs the "management by exception" principle. In evaluating the forecast accuracy of such systems, a cost-effectiveness approach must be taken. A forecast model which performs well in one application may be totally unsuited in another

where the forecast accuracy required is more stringent. Since any forecasting system selected for use must be satisfactorily adaptable to a wide range of demand patterns it is of interest to examine a few of the short-term forecasting models which are currently in vogue.

This thesis investigates the types of time series for which various forecast models are appropriate and the degree to which their forecast performance is degraded by changes in the forecast series generation model. The results of this investigation should suggest which model or models tend to be most adaptive in the sense that satisfactory forecasts are consistently made for the particular series forecast in the study. It is realized that forecast acceptability is situation-dependent, and possibly somewhat subjective. Also, the results obtained will necessarily be highly dependent on the series upon which forecasts are based, and the series used in this study may not be representative of the many demand or failure patterns which must be forecast in practice. One must also remember that the optimality (minimum variance unbiasedness) of even the least versatile forecast model may be demonstrated. As Bossons [Ref. 13] has concluded:

(a) if it can be shown that the model corresponds to a linear transformation of the stochastic process assumed to generate the time series to be forecast by the model, and

(b) if efficient estimates can be derived for the parameters of the stochastic model and thus also of the rule for "adapting" the forecasts, then the optimality of <u>any</u> rule can be demonstrated.

The decision to trade this possibly restricted optimality for versatility can only be made in the environment where the model is to be ultimately used. It should also be noted that the measure of optimality associated with the EWMA models differs from that used in the least-squares models and for this reason a brief discussion of the difference is needed. The EWMA models use the method of discounted least squares (D.L.S.). An excellent discussion by Gilchrist of this and other discounting methods is contained in Reference 15. His modification of the method of moments, least-squares and maximum likelihood to incorporate discounting may offer a fertile area for future investigation in which the modified methods are compared to the traditional procedures. He argues that the discounted estimators have more robust properties than the standard methods.

The method of discounted least squares is normally used with models of the general form

$$Y_t = f(t; X_{t-1}, \ldots, X_o; \theta) + \varepsilon_t, \qquad t=0,1,\ldots$$

where the t^{th} observation is expressed as a function of all previous observations, a parameter θ, and time t. The estimates of θ are chosen to minimize the discounted sum of squared errors

$$\sum_{i=0}^{t} W_t e_{t-i}^2$$

where W_t is usually of the form $\beta^t, 0<\beta<1$. In using this form of W_t, a weighted average results which gives maximum weight to the most recent data. The estimators are easily updated by simple recursive relations as each new observation is taken. The comparisons made in this investigation were based only on the mean squared error of forecast, however, and no discounting was involved. It should be made clear that forecast performance has been measured objectively without regard to the type of optimality which each model seeks to obtain. There is no attempt in this thesis to extend conclusions beyond such inherent characteristics. However, it is hoped that subsequently, when operational time series data must be forecast using a model, the comparison methodology expounded here will aid in the selection of the "best" model for that data.

II. SHORT-TERM FORECASTING MODELS

A. LEAST-SQUARES ESTIMATION MODELS

This classical prediction method has evolved, using regression analysis, by considering a functional relationship between the observed values of the random process and an independent or control variable. The General Linear Hypothesis is widely known and therefore will not receive more than a brief statement in this section. An excellent discussion of the conditions under which this approach is most appropriate may be found in Reference 1.

Zehna [Ref. 1] and Coventry [Ref.2] have investigated the advantages of maximum likelihood estimation procedures and have shown that these methods are superior to exponential smoothing methods for both the constant mean case and the linear model, where the forecast errors are assumed to be normally distributed in both cases. No comparisons were conducted for non-normal, or auto-correlated series in which the maximum likelihood assumptions are not appropriate. The relative robustness of the least squares method for various series generation model specifications is one of the objectives of this investigation. The simple regression model also provides a standard reference point from which the performance of exponentially weighted forecast models may be measured.

1. Simple Least-Squares Forecasting

a. Series Generation Model

The process observations are assumed to be a linear function of time, the independent variable. The model may be written

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \tag{A-1}$$

where Y_t is the random process observation taken at time X_t, and ε_t is the random component contained in that observation. It is assumed that $E[\varepsilon_t] = 0$ and $Var(\varepsilon_t) = \sigma^2$ for all t. For this investigation β_1 is taken to be zero in the generation model, and all observations are relative to that level.

Consideration of the generation model reveals that forecast errors will be composed of two components, errors of parameter estimation and the random variation from the true linear model, Equation (A-1). As the number of observations increases, it may be observed that the parameter estimates will exhibit smaller variance until the forecast error is almost entirely composed of the random variation inherent in the process.

b. Parameter Estimation

The least squares estimate of β_2 in the zero-intercept form based on observations $y_1, y_2, \ldots, y_n$ is given by

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} . \tag{A-2}$$

This form is used since all series in this comparison begin at the origin, $X_o = 0$, $Y_o = 0$. The reader unfamiliar with the development of this estimator is referred to References 1 and 2. It may be observed that Equation (A-2) above is also the maximum likelihood estimator of β_2 when it is further assumed that ε_t has a normal distribution. The variance of the estimated value of the process is given by

$$\mathrm{VAR}(\hat{Y}_t) = \sigma_\varepsilon^2 \frac{X_t^2}{\sum_{i=1}^{t} (X_i - \bar{X})^2}$$

where σ_ε^2 is the (constant) variance of the random component.

c. The Forecast Model

After parameter estimation using past observations through time X_{t-1}, the forecast value of the random series at time X_t is given by

$$F_t = \hat{Y}_t = \hat{\beta}_2 X_t. \tag{A-3}$$

Although, as noted in Reference 1, the regression line must not be extrapolated too far beyond the range of X_t values this method is well suited for one-step-ahead prediction. In fact, it may be shown [Ref. 1] that Equation (A-2) is the zero intercept minimum variance unbiased linear estimator for β_2.

2. Modified Least-Squares Estimation

a. Consequences of Autocorrelated Disturbances

One of the crucial assumptions when dealing with least-squares estimation models is the serial independence

of the disturbance term which is implied in $E[\varepsilon\varepsilon'] = \sigma^2 I$ and gives $E[Y_t \; Y_{t+s}] = 0$ for all $s \neq 0$. Autocorrelated disturbances arise frequently in the estimation of relationships from time series data, and it will be seen that almost all the series used in this comparison of forecast models have significant autocorrelation. Some mention of the main consequences of autocorrelated disturbances is therefore in order. In the presence of such disturbances, the estimates of β_1 and β_2 obtained by simple least-squares estimation remain unbiased, but the variances of the estimator may be large. This suggests the possibility of modifying the procedure to reduce the variance. Further, if the usual least-squares methods are applied, an underestimate of these estimator variances is likely to occur.

b. Series Generation Model

The same linear model will be assumed here as before in Equation (A-1) $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$ except now it will be assumed that the disturbances or shocks are no longer independent random variables. Instead, a first-order autoregressive model is used which is given by

$$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t \qquad \text{(A-4)}$$

where $\{\delta_t\}$ has mean zero, variance σ_δ^2 and zero covariances. To give a better basis for comparison of the maximum likelihood or least-squares model with the modified model, the same normal variates will be used to generate the sequence $\{\delta_t\}$ in Equation (A-4) above as were used to generate the sequence $\{\varepsilon_t\}$ in Equation (A-1).

c. The Modified Forecast Model

To predict Y_{t+1} for a given value of X_{t+1} the conditional expected value of Equation (A-1) may be taken, giving

$$E[Y_{t+1} | \varepsilon_1, \ldots, \varepsilon_t] = \beta_1 + \beta_2 X + E[\varepsilon_{t+1} | \varepsilon_1, \ldots, \varepsilon_t]$$

$$= \beta_1 + \beta_2 X_{t+1} + \rho\varepsilon_t .$$

Here the result involves the conditional expectation of Equation (A-4) also. Substituting the value of ε_t from Equation (A-4) gives, after rearrangement,

$$E[Y_{t+1} | \varepsilon_1, \ldots, \varepsilon_t] = \beta_1(1-\rho) + \beta_2(X_{t+1} - \rho X_2) + \rho Y_t .$$

This can be rewritten as

$$E[Y_{t+1} - \rho Y_t | \varepsilon_1, \ldots, \varepsilon_t] = \beta_1(1-\rho) + \beta_2(X_{t+1} - \rho X_t) .$$

If ρ is known, Equation (A-1) may then be written

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t + \rho X_{t-1}) + \delta_t . \qquad \text{(A-5)}$$

This transformation satisfies in full the assumptions of the simple linear model, and permits the direct application of least squares to the transformed variables $(Y_t - \rho Y_{t-1})$ and $(X_t - \rho X_{t-1})$ yielding the best linear predictors. The best linear predictor of $(Y_{t+1} - \rho Y_t)$ is $\hat{\beta}_1(1-\rho) + \hat{\beta}_2(X_{t+1} - \rho X_t)$ where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the least-squares estimators of the parameters in Equation (A-5). This is equivalent to saying that the best linear predictor of Y_{t+1}, conditional on past observations, is $\hat{\beta}_1(1-\rho) + \hat{\beta}_2(X_{t+1} - \rho X_t) + \rho Y_t$ which may be written more clearly as

$$\hat{Y}_{t+1} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+1} + \rho[Y_t - (\hat{\beta}_1 + \hat{\beta}_2 X_t)] . \qquad \text{(A-6)}$$

Analysis of this form of the predictor provides some insight and permits comparison with the unmodified form of the model. By incorporating the available knowledge of the autoregressive structure of the process a term is added to the estimate of Y_{t+1}. This term is the product of ρ and the estimated disturbance (actually the forecast error) of the previous period.

For a further discussion of the modified model and the effect of autocorrelation on the linear model, Chapter 7 of Reference 12 is recommended.

B. THE SIMPLE EXPONENTIALLY WEIGHTED MOVING AVERAGE FORECAST MODEL

1. Integrated Moving Average (IMA) Processes

Since there is often sufficient reason to believe that the more recent observations of a process represent that process better than older data, EWMA methods which give larger weight to the more recent data in constructing and adjusting the model have been developed. It is intuitively clear that such discounting of older data is not applicable when a stationary process is observed, since even the oldest observations are as representative of the process as the most recent.

Stochastic processes of the type for which EWMA forecasts are appropriate have been described by Box and Jenkins [Ref. 3] as integrated moving average processes. These processes act as though no fixed mean exists, but do exhibit homogeneity in the sense that, apart from local

level and trend, one part of the series behaves much like any other part. To clarify the description of IMA processes three equivalent forms of the model may be formulated [Ref. 3].

a. The difference equation form may be written

$$\nabla Y_t = (1-\beta B)\varepsilon_t \qquad -1<\alpha<1$$

where ∇ is the difference operator and B is the backwards operator. The model may also be written using these definitions as

$$Y_t = Y_{t-1} - \beta\varepsilon_{t-1} + \varepsilon_t . \tag{B-1}$$

b. The random shock form follows from the difference equation form by use of recursive substitution for the older observations,

$$Y_t = \alpha \sum_{j=1}^{t-1} \varepsilon_j + \varepsilon_t \tag{B-2}$$

where $y_o = 0$.

c. The inverted form follows, writing recursively in terms of the observations rather than the random shocks,

$$Y_t = \hat{Y}_t + \varepsilon_t$$

where

$$\hat{Y}_{t-1} = \alpha \sum_{j=1}^{\infty} \beta^{j-1} Y_{t-j}$$

from which form, the model can easily be written in recursive form

$$\hat{Y}_{t+1} = \alpha Y_t + \beta\hat{Y}_t . \tag{B-3}$$

This is recognizable as the recursion formula for the exponentially weighted moving average. It follows then that EWMA forecasts should be appropriate for IMA processes. A demonstration that this is true is contained in section B.3. The form of Equation (B-2) is most convenient for generating IMA series, while the form of Equation (B-3) is most often used for forecasting.

2. The Series Generation Model

The specific EWMA generating model used in this study can be treated as a simple random walk process, one which is reasonably stable with no trends. The model is given by

$$\begin{aligned} Y_t &= \bar{Y}_t + \varepsilon_t \\ \bar{Y}_t &= \bar{Y}_{t-1} + \gamma_t \end{aligned} \qquad \text{(B-4)}$$

where Y_t is the observed value, $\bar{Y}_t$ is the true level of the process at time t, γ_t is a slight random change from the prior process level, and ε_t is a random shock or superimposed error. ε_t and γ_t are uncorrelated random variables with zero mean and constant variances $V(\varepsilon_t)$ and $V(\gamma)$ respectively. It will be shown in section B.5 that the EWMA forecast model supplies the optimal linear least-squares predictor for this generating model. A further result, provided $0<\alpha\leq 1$, is that any generating process for which the EWMA is optimal can be represented by the generating model above in Equation (B-4) even though this model may not necessarily describe the true generating process. It must, however, describe the external behavior of the true generating process accurately

or else the EWMA would not then be optimal. Stated another way, if a process cannot be represented by Equation (B-4) then the EWMA forecast cannot be an optimal forecast rule for that series.

3. Optimal Properties of Exponentially Weighted Forecasts

It has been suggested intuitively in the preceding discussion of IMA processes that the EWMA recursion relation describes the process exactly. Following the approach of Muth [Ref. 4], it can be shown that for an IMA process the EWMA equals the conditional expected value of the process.

The EWMA forecast results from a model of expectations which are adapted to the most recent information concerning the process. Let Y_t represent that part of a time series which cannot be explained by systematic factors such as seasons or trends in the average. Let $\hat{Y}_t$ represent the forecast, or conditional expectation of Y_t which is made at time t-1 on the basis of available information at that time. Assume that the forecast is changed from one period to the next by an amount proportional to the latest observed error. This implies that a permanent component exists in every observed error and that the level of the process subsequently reflects this component:

$$\hat{Y}_t = \hat{Y}_{t-1} + \alpha(Y_{t-1} - \hat{Y}_{t-1}) \quad 0 \le \alpha \le 1 \qquad \text{(B-5)}$$

This in turn yields the EWMA formula:

$$\hat{Y}_t = \alpha \sum_{i=1}^{\infty} \beta^{i-1} Y_{t-i}. \qquad \text{(B-6)}$$

Since $\beta = 1-\alpha$, the weights corresponding to previous values of Y_t do not introduce any systematic bias.

Assume now that the observed value of the process can be written as a linear function of the independent random shocks:

$$Y_t = \sum_{i=1}^{\infty} W_i \, \varepsilon_{t-i} + \varepsilon_t \tag{B-7}$$

where the shocks are i.i.d. with mean zero and variance σ^2.

If the parameters W_i which characterize the random process are known, the expected value of Y_t may be easily found. If it is desired to do so one period in advance of time t when ε_t is as yet unknown, the conditional expected value of Y_t given the past values ε_{t-1}, ε_{t-2}, ..., may be used to yield

$$\hat{Y}_t = E[Y_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots] = \sum_{i=1}^{\infty} W_i \, \varepsilon_{t-i}. \tag{B-8}$$

To relate the regression function above to the EWMA expression, Equation (B-8) must be written as follows in terms of the observations, where the values of the coefficients υ_j must be determined.

$$\hat{Y}_t = \sum_{j=1}^{\infty} \upsilon_j \, Y_{t-j}. \tag{B-9}$$

Substituting for Y_{t-j} with the random shock form and rearranging terms results in:

$$\hat{Y}_t = \sum_{j=1}^{\infty} \upsilon_j \left({}_{t-j} + \sum_{i=1}^{\infty} W_i \; {}_{t-i-j} \right) = \upsilon_1 \, \varepsilon_{t-1} + \sum_{i=2}^{\infty} \left(\upsilon_i + \sum_{j=1}^{i-1} \upsilon_j \, W_{i-j} \right) \varepsilon_{t-i} . \tag{B-10}$$

A relationship between the parameters (W_i) associated with the unobserved shocks and the coefficients (υ_j) associated with the past observations of the process, is obtained by comparing coefficients of their respective expressions

$$W_1 = \upsilon_1$$

$$W_i = \upsilon_i + \sum_{j=1}^{i-1} \upsilon_j W_{i-j} \quad i = 2,3,4,\ldots$$

To demonstrate that the EWMA forecast evolves from the above process, the weights $\upsilon_j = \alpha\beta^{j-1}$, $j = 1,2,3,\ldots$ from the inverted form are substituted into the above expression. This results in the system:

$$W_1 = \alpha$$

$$W_i = \alpha\,\beta^{i-1} + \alpha \sum_{j=1}^{i-1} \beta^{j-1} W_{i-j}, \quad i = 2,3,4,\ldots \tag{B-11}$$

It can be seen from the above that $W_i = \alpha$ for all $i \geq 1$.

Writing Y_t in the random shock form and substituting $\alpha = W_i$ results in a form comparable to Equation (B-2)

$$Y_t = \alpha \sum_{i=1}^{t-1} \varepsilon_{t-i} + \varepsilon_t.$$

Consideration of this expression reveals that the shock associated with each time period has a weight of unity, but previous shock weights are constant with a weight between zero and one. This demonstrates that part of the random shock experienced in a period has a permanent effect, but the rest of the shock affects the process only during the current period.

The foregoing assumed that forecasts were always for one period ahead only, but it can be shown that the best forecasts for all future periods are the same. The proof may be found in [Ref. 4] and is a generalization of the above approach where $Y_{t,T} = \sum_{i=0}^{\infty} W_{i+T} \varepsilon_{t-i}$ represents the forecast T periods ahead. The unique solution for the coefficients results in $\upsilon_{T,k} = \alpha \beta^{k}$, $k = 0,1,2,\ldots$ independent of T as asserted. The reason for this result is that all prior shocks are weighted equally and the forecasts then only estimate the permanent component of the shocks.

Muth [Ref. 4] also shows that the EWMA forecast rule is appropriate if the permanent and transitory components are independent rather than perfectly correlated as assumed above. The smoothing constant in the independent case however is constrained to be a function of a characteristic root of a system of difference equations.

4. The Forecast Model as an IMA Process Generator

The familiar EWMA forecast model is identical to the IMA process in Equation (B-3) $F_{t+1} = \hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$, which may be rewritten in the random shock or forecast error form of Equation (B-2) by substitution of the error $\varepsilon_t = Y_t - \hat{Y}_t$. This yields

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha \varepsilon_t .$$

Successive substitutions for the oldest estimate in terms of the prior estimate plus the permanent component of the random shock yields a form equivalent to Equation (B-2)

$\hat{Y}_{t+1} = \hat{Y}_o + \alpha \sum_{i=0}^{t} \varepsilon_i$, and here $\hat{Y}_o = 0$ is chosen level from which the process is measured. The form of Equation (B-2) is such that a series of autocorrelated forecasts may be easily generated. Following Muth's interpretation [Ref. 4], this generation method merely adds the permanent component of the random shock to the current process estimate to yield the next period estimate.

Since the next period has an associated forecast error, that error is then added to produce the desired observation:

$$Y_{t+1} = F_{t+1} + \varepsilon_{t+1} = \alpha \sum_{i=0}^{t} \varepsilon_i + \varepsilon_{t+1}$$

which by substitution of Equation (B-2) is in a form suitable for use in the Fortran series generation model. This model, having no fixed mean, is free to drift in a random walk. It may be observed that when ε_t is reduced to zero, the model no longer experiences shocks and becomes deterministic at the last $\hat{Y}_t$ for which $\varepsilon_t > 0$. The asymptotic variance of the generated series may be determined by observing from Equation (B-2) with $t \to \infty$ that $E[Y] = E[\varepsilon] = 0$ and

$$\sigma_y^2 = E[Y^2] - E[Y]^2 = E\left[\alpha \sum_{i=0}^{\infty} \beta^i \, Y_{t-i}\right]^2 = \frac{\alpha}{2-\alpha}\,\sigma^2. \quad \text{(B-12)}$$

In generating the autocorrelated series, the constant α controls both the amount of correlation which exists and the series variance. When $\alpha=1$, the series is perfectly correlated with the random series and has variance σ^2.

5. Parameter Estimation Procedures

To develop optimal parameters for a specific series, consider a strictly stationary stochastic process with

$$E[Y_t] = 0, \; VAR[Y_t] = \sigma^2, \; COV[Y_t, Y_{t+k}] = \rho_k \sigma^2 .$$

If the EWMA model is used for forecasts, then the mean squared error of prediction is by definition:

$$M.S.E. = E[Y_t - \hat{Y}_t]^2 .$$

Expanding this expression,

$$M.S.E. = E[Y_t^2] - 2E[Y_t \hat{Y}_t] + E[\hat{Y}_t^2] \tag{B-13}$$

where $E[Y_t^2] = \sigma^2$. (B-14)

Substituting for $\hat{Y}_t$ in the second term above gives

$$-2E[Y_t \hat{Y}_t] = -2E\left[Y_t \; \alpha \sum_{i=0}^{\infty} \beta^i \; Y_{t-i}\right]$$

$$= -2\alpha \sum_{i=0}^{\infty} \beta^i \; E[Y_t \; Y_{t-i}] \tag{B-15}$$

$$= -2\alpha \sum_{i=0}^{\infty} \beta^i \; \rho_i \; \sigma^2 = -2\alpha\sigma^2 \sum_{i=0}^{\infty} \beta^i \; \rho_i \; .$$

Expanding the third term in Equation (B-13) and substituting for $\hat{Y}_t$ yields

$$\alpha^2 E\left[\sum_{i=0}^{\infty} \beta^i \; Y_{t-i}\right]^2 .$$

Expanding the expression formally results in

$$\alpha^2 E\left[Y_t + \beta \; Y_{t-1} + \beta^2 \; Y_{t-2} + \cdots\right]^2 = \alpha^2 W \tag{B-16}$$

where $W = E[Y_t + \beta Y_{t-1} + \beta^2 Y_{t-2} + \cdots]^2$. (B-17)

Since $\{Y_t\}$ is a strictly stationary process, W may also be written $W = E[Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + \cdots]^2$. Rearranging W this becomes

$$W = E[Y_t + \beta(Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + \cdots)]^2$$

which may now be expanded. Proceeding formally, this procedure results in

$$W = E[Y_t^2] + 2E[Y_t \beta(Y_{t-1} + \beta Y_{t-2} + \beta^2 Y_{t-3} + \cdots)] + \beta^2 W$$

which may be written more briefly as

$$W = \sigma^2 + 2E\left[\sum_{i=1}^{\infty} \beta^i Y_t \, Y_{t-i}\right] + \beta^2 W$$

$$= \sigma^2 + 2\sigma^2 \sum_{i=1}^{\infty} \beta^i \rho_i + \beta^2 W = u + \beta^2 W$$

Here u is defined as the sum of the first two terms. Solving for W, this becomes $W = \frac{u}{1-\beta^2}$. Again substituting W into Equation (B-16) results in

$$\alpha^2 W = \frac{\alpha u}{1+\beta} \quad \text{(B-18)}$$

Combining Equation (B-14), Equation (B-15), and Equation (B-18) in Equation (B-13) yields

$$\text{M.S.E.} = \frac{2\delta^2}{2-\alpha} - \frac{2\alpha\sigma^2}{2-\alpha} \sum_{i=1}^{\infty} \beta^{i-1} \rho_i \quad \text{(B-19)}$$

which is the mean squared error of one-step-ahead prediction as a function of the smoothing constant and the autocorrelation of the series. This is a result obtained by Cox, p. 415, Ref. 5. For a Markov series with exponential autocorrelation

$$\rho_k = \rho^k \ , \ k \geq 0$$

where k is the lag between observations, the mean squared error for one-period ahead forecasts becomes

$$\text{M.S.E.} = \frac{2\sigma^2(1-\rho)}{(1-\beta\rho)(1+\beta)} \tag{B-20}$$

The MSE is minimized for a given ρ by an EWMA predictor which has parameters

$$\beta_{opt} = \begin{cases} \frac{1-\rho}{2\rho} \\ \\ 1 \end{cases} \text{or } \alpha_{opt} = \begin{cases} \frac{3\rho-1}{2\rho} & (\frac{1}{3} \leq \rho \leq 1) \\ \\ 1 & (-1 \leq \rho \leq 1/3) \end{cases} \tag{B-21}$$

These results are obtained by equating to zero the derivative of Equation (B-20). The corresponding mean squared error of prediction for the EWMA with optimal parameters is obtained from Equation (B-20) by substitution

$$\text{M.S.E.} = \begin{cases} \frac{8\sigma^2\rho(1-\rho)}{(1+\rho)^2} \ , & (1/3 < \rho \leq 1) \\ \\ \sigma^2 \ , & (-1 \leq \rho \leq 1/3) \end{cases}$$

Analysis of this result discloses that when $\rho > 1/3$ it is optimal to predict the observed series values using a larger α value since there is sufficient correlation between successive observations to make this procedure advantageous. When $\rho \leq 1/3$ the EWMA should attempt to predict only the mean value of the process since so much uncertainty exists concerning the next observation. When the optimal α equals zero, the MSE of prediction equals the variance of the time

series. When $\rho=\alpha=1$ perfect correlation permits exact forecasts and the M.S.E. is zero.

The preceding results are obtained by Cox [Ref.5] who also developed tables showing the relationships between ρ, α_{opt}, and the MSE/σ^2 ratio. It is interesting to note that as one predicts more than one step ahead, the critical value of ρ increases. One conclusion drawn from Cox's tables is that MSE is not sensitive to the choice of a smoothing constant, and although the optimal α may be zero, it costs little to use $\alpha = 0.1$ as insurance against a possible change in the mean level of the process. This insensitiveness probably accounts for the success of the exponential smoothing model proposed by Brown [Ref. 6] for which he uses empirical methods of parameter estimation or a rule of thumb, $\alpha=0.1$ for general use.

An alternative approach to optimal parameter derivation for the EWMA forecast model used by Harrison [Ref. 7] may be taken by considering the generating model in Equation (B-4). Recalling that the one step ahead forecast error $e_t = Y_t - F_t = Y_t - \hat{Y}_t$, and that the forecast model is $\hat{Y}_t = \hat{Y}_{t-1} + \alpha e_t$, a difference equation may be written to express the one step ahead forecast error

$$e_{t+1} = (1-\alpha)e_t + \varepsilon_{t+1} - \varepsilon_t + \gamma_{t+1}. \qquad \text{(B-22)}$$

Then the expectation of the product of Equation (B-22) with both e_{t+1} and e_t yields the variance and first covariance respectively of the one-step-ahead forecast errors

$$V = (1-\alpha)C + (1+\alpha)\ V(\varepsilon) + V(\gamma) \tag{B-23}$$

$$C = (1-\alpha)V - V(\varepsilon).$$

Solving these two equations for V yields

$$V = \frac{2\alpha\ V(\varepsilon) + V(\gamma)}{\alpha(2-\alpha)} . \tag{B-24}$$

Differentiating this with respect to α gives the minimum variance parameter

$$\alpha = \frac{(1+4R)^{\frac{1}{2}} - 1}{2R} \quad \text{where } R = \frac{V(\varepsilon)}{V(\gamma)} . \tag{B-25}$$

Substituting for α in Equation (B-24) it is found that the minimum variance is $V = \frac{V(\varepsilon)}{1-\alpha}$ and the covariance is 0. The error covariances may easily be shown to be zero by taking the expectation of the product of Equation (B-22) with ε_{t-i} with $i \geq 2$. All covariances are zero, so it follows that the predictor is optimal. Although this approach was used by Harrison [Ref. 7], he then recommended that because of the usual shortage of data, general robustness, and precision, for short-term forecasting the optimal parameter should be derived from a simulation of the predictor on the original series data. This has also been the suggestion of Cox, Brown and others. It appears that even though the optimal parameters may be determined for particular series on a theoretical basis, it is faster and possibly about as satisfactory in terms of forecast results, to determine the parameters empirically. The insensitiveness of the simple EWMA model to the parameter value chosen has made parameter selection an area of rapidly diminishing return for the effort expended.

C. COX'S MODIFIED EXPONENTIALLY WEIGHTED MOVING AVERAGE FORECAST MODEL

1. Series Generation Model for the Exponential Autocorrelation Function

Cox [Ref. 5] has shown, as noted in Section B.4, that the exponentially weighted moving average is an optimal forecast method for time series having an autocorrelation coefficient with lag k of the form $\rho(k) = \beta^k$. A recursion relation which is identical to the relation for exponential smoothing and which generates such a time series is shown by Naylor [Ref. 8]. This recursion relation is given by

$$Y_o = (1-\beta)\varepsilon_o$$

$$Y_t = \beta Y_{t-1} + (1-\beta)\varepsilon_t$$

where $\{\varepsilon_t\}$ are mutually independent variables with zero mean and variance σ_ε^2.

The method used in this investigation to produce the exponential autocorrelation function requires that the ε_t be uniform random numbers in the interval [-A,A]. This choice of ε_t results in exponentially autocorrelated variates Y_t which have zero mean and variance equal to

$$\sigma_y^2 = \frac{1-\beta}{1+\beta}\sigma_\varepsilon^2 = \frac{\alpha}{2-\alpha}\sigma_\varepsilon^2$$

where α, the smoothing constant, equals $1-\beta$. That the recursion relation above gives the desired result may be seen by using recursive substitution so that $Y_t = \alpha \sum_{j=0}^{t-1} \beta^j \varepsilon_{t-j} + \beta^t Y_o$ or in terms of the random variable alone

$$Y_t = \alpha \sum_{j=0}^{t-1} \beta^j \varepsilon_{t-j} + \alpha\beta^t \varepsilon_o.$$

The mean of the generated time series follows immediately since $E[Y] = E[\varepsilon] = 0$ and the variance is

$$\sigma_y^2 = E[Y^2] - E[Y]^2 = E[Y^2].$$

Squaring the recursion relation and taking expected value gives $\sigma_y^2 = \beta^2 E[Y_{t-1}^2] + 2\alpha\beta E[\varepsilon_t Y_{t-1}] + \alpha^2 E[\varepsilon_t^2]$ which gives, since ε_t and Y_{t-1} are independent with mean zero, the result

$$(1-\beta^2)\sigma_y^2 = \alpha^2\sigma_\varepsilon^2.$$

Therefore, $\sigma_y^2 = \frac{\alpha}{1+\beta}\,\sigma_\varepsilon^2$, which is equivalent to the variance expression above. When, as in this investigation, the ε_t are uniformly distributed on the range $[-A,A]$ the variance becomes

$$\sigma_y^2 = \frac{\alpha}{1+\beta}\;\frac{(2A)^2}{12} = \frac{\alpha A^2}{3(1+\beta)}$$

The autocorrelation for two numbers which are k observations apart in the time series is

$$\phi(k) = E[Y_t Y_{t+k}] - E[Y_t]E[Y_{t+k}] = \frac{\alpha}{2-\alpha}\;\beta^k\sigma_\varepsilon^2$$

which implies that the autocorrelation coefficient is $\rho(k) = \frac{\phi(k)}{\sigma_y^2} = \beta^k$ which is the desired result. This result is obtained by making a transformation on the summations in the product $Y_t Y_{t+k}$ which makes use of the stationarity of the series ε_t and then taking the expected value of the product $Y_t Y_{t+k}$ while both are expressed in terms of the random shocks. The transformation referred to above is

$$Y_{t+h} = \alpha \sum_{j=0}^{t+k+1} \beta^j \varepsilon_{t+k-j} + \alpha\beta^{t+k}\varepsilon_0 = \alpha \sum_{j=0}^{t-1} \beta^{j+k}\varepsilon_{t+k-(j+k)}$$

$$+ \alpha\beta^{t+k}\varepsilon_{-k} = \beta^k \left[\alpha \sum_{j=0}^{t-1} \beta^j \varepsilon_{t-j} + \alpha\beta^t \varepsilon_{-k} \right]$$

which has the same expectation as $\beta^k E[Y_t]$. Therefore, $E[Y_t \; Y_{t+h}] = \beta^k \; E[Y_t^2] = \beta^k \sigma_y^2 = \phi(k)$ from which the result follows immediately.

2. Wiener's Optimal Linear Predictor

Cox [Ref. 5], after developing the optimal smoothing parameters for exponentially correlated ($\rho_k = \rho^k$) time series which are shown in Section B.4, sought to improve the model so that it did not lag behind when used to forecast series containing an increasing mean value. He made use of Wiener's optimal linear predictor for this same series which is given by:

$$\hat{Y}_t(\text{opt.}) = \rho^h(Y_t - \mu) + \mu = \rho^h \; Y_t + (1-\rho^h)\mu \tag{C-1}$$

where h is the number of periods ahead for which a forecast is desired, and ρ is the autocorrelation coefficient of the series to be forecast. The Wiener predictor has a mean square error of forecast which may be written

$$\text{M.S.E.}_W = (1-\rho^{2h})\sigma_\varepsilon^2.$$

The Wiener predictor, although shown optimal when all assumptions were met, had two properties which made it objectionable for practical use. First, the parameters μ and ρ are assumed known. This can be overcome if sufficient past observations are available for use in estimating μ and ρ, since the mean squared error is asymptotically unaffected by the substitution of unbiased estimates for these parameters. The more serious objection to the predictor, Equation

(C-1) however, is that it becomes biased if the mean shifts, correspondingly increasing the mean square error of forecast.

3. The Modified Form of the Model

Cox extended Equation (C-1) by substituting $\hat{Y}_t$ for μ so that the predictor would follow shifts in the process mean. Equation (C-1) then becomes

$$\hat{Y}_m = \rho^h Y_t + (1-\rho^h)\hat{Y}_t \tag{C-3}$$

or substituting for $\hat{Y}_t$, it may be written

$$\hat{Y}_m = (\alpha+\beta\rho^h) Y_t + \beta (1-\rho^h)\hat{Y}_{t-1}. \tag{C-4}$$

Inspection of these two equations reveals that as $\alpha \to 0$ the MSE of Equation (C-3) approaches that of the Wiener predictor, Equation (C-2), since Equation (C-4) collapses to Equation (C-1). Also note that when $\rho \neq 0$ the model Equation (C-3) is no longer an exponentially weighted moving average. It is instead a moving average in which the weights of past observations decrease exponentially, but in which the current observation receives a weight which is not a member of the same geometric series. The weights appearing in the expression for $\hat{Y}_t$ in Equation (B-3) sum to one, as do the coefficients of Y_t and $\hat{Y}_t$ in Equation (C-3), thereby maintaining the desired average. Another observation concerning Equation (C-4) is that as the number of periods ahead for which we wish to forecast increase, the modified predictor tends to reduce to the simple EWMA form.

4. Parameter Estimation

The mean squared error of Equation (C-3) is given by

$$\text{M.S.E.} = \frac{2\sigma_\varepsilon^2(1-\rho^h)(1-\rho\beta^2+\beta\rho^h-\beta\rho^{h+1})}{(1-\beta\rho)(1+\beta)} \qquad \text{(C-5)}$$

which reduces to the mean squared error of the Wiener predictor Equation (C-2) when $\beta=1=h$. The values selected for α and β in Equation (C-4) will necessarily reflect a compromise between the desire for a minimum mean square error and a desire for protection against a change in mean process value. As Cox has shown [Ref. 5] however, when the optimal value of α is 0 there is insignificant change in mean squared forecast error for the EWMA when α is made larger up to the range 0.1 or 0.15.

D. THE HOLT-WINTERS AND THEIL-WAGE TWO-PARAMETER EXPONENTIALLY WEIGHTED MOVING AVERAGE (EWMA) LINEAR GROWTH FORECAST MODELS

Holt [Ref. 9] as further discussed in Winters [Ref. 10] proposed a forecasting model for time series exhibiting linear growth which was a simple extension of the EWMA method already applied to time series with constant mean. Although Winters' development explained precisely how the method worked, he has been criticized by Theil and Wage [Ref. 11] and others for failing to explicitly justify the method. The rationale for the criticism seems to be that Winters failed to formulate an explicit stochastic model as a basis for the forecasting method. Instead, Winters used a completely empirical method for selection of parameters. It

must be observed, however, that Winters utilized time series of actual data from three dissimilar processes. His approach is probably representative of those currently used when real data must be forecast. For real data, when the model is in fact unknown, Winters' approach implicitly assumes various underlying models and his choice of parameters which gave the least forecast variance is his implicit specification of the underlying model. For the purpose of this investigation the criticism raised by Theil and Wage is considered valid, and their generating model will be used along with a discussion of their determination of optimal parameters. It may be added that Harrison [Ref. 7] has postulated yet another generating model for the Holt-Winters predictor which differs slightly from the Theil-Wage model and has also derived optimal parameters.

1. <u>The Linear Growth Series Generation Model</u>

The Linear growth model proposed by Theil and Wage [Ref. 11] is given by

$$\begin{aligned} Y_t &= \bar{Y}_t + \varepsilon_t \\ \bar{Y}_t &= \bar{Y}_{t-1} + b_t \qquad\qquad \text{(D-1)} \\ b_t &= b_{t-1} + \delta_t \ . \end{aligned}$$

Here ε_t is again a random shock or residual by which the observation Y_t differs from its mean value $\bar{Y}_t$ and the change attributed to trend or the slope b_t is changed by a random amount δ_t. The random variables ε_t and δ_t are assumed to have zero mean, constant variance and zero covariances.

2. The Forecast Models

a. The Holt-Winters Model

The Holt-Winters forecast model is given by

$$F_t = \hat{Y}_t + k\hat{b}_t \tag{D-2}$$

where k = 1 for one-step-ahead forecasts. This is the original exponential smoothing model in which

$$\hat{Y}_t = \alpha\, Y_t + \beta(\hat{Y}_{t-1} + \hat{b}_{t-1}) \tag{D-3}$$

and

$$\hat{b}_t = \alpha'\, (\hat{Y}_t - \hat{Y}_{t-1}) + \beta'\hat{b}_{t-1}. \tag{D-4}$$

The first parameter α in Equation (D-3) is the smoothing constant for the process level. Note that Equation (D-3) is identical to (B-3) except for the addition of the slope to the previous process level. The second parameter α' in Equation (D-4) is the smoothing constant for the exponentially smoothed slope estimate. Holt and Winters place no constraints on the relation between α and α' as do Brown [Ref. 6] and Theil and Wage [Ref. 11].

b. The Theil-Wage Model

Theil and Wage have suggested that Equation (D-3) neglects available information and propose to replace it with

$$\hat{Y}_t = \alpha\, Y_t + \beta(\hat{Y}_{t-1} + \hat{b}_t) \tag{D-5}$$

which they consider a more simultaneous, rather than a recursive, approach. Taking Equation (D-4) and substituting into (D-5) they obtain

$$\hat{Y}_t = \alpha\, Y_t + \beta[\hat{Y}_{t-1} + \alpha'(\hat{Y}_t - \hat{Y}_{t-1}) + \beta'\hat{b}_{t-1}]$$

$$= \beta\alpha'\hat{Y}_t + \alpha Y_t + \beta\beta'(\hat{Y}_{t-1} + \hat{b}_{t-1})$$

and therefore, solving for $\hat{Y}_t$

$$\hat{Y}_t = \frac{\alpha}{1-\beta\alpha'}\, Y_t + \frac{\beta\beta'}{1-\beta\alpha'}\,[\hat{Y}_{t-1} + \hat{b}_{t-1}] \qquad \text{(D-6)}$$

which is of the same general form as Equation (D-3), but the smoothing and discount parameters are changed to obtain all the information contained in the observations. Note that when $\alpha' = 0$, $\beta' = 1$. Equation (D-6) becomes identical to Equation (D-3), because when $\alpha' = 0$, correspondingly, the slope is constant and therefore makes no contribution.

3. Parameter Estimation Procedures

a. The Holt Winters Model

The empirical parameter selection method used by Winters [Ref. 10] was a steepest ascent search for the optimal parameters. While this may be appropriate for critical forecasts, it is recalled that the objective of this comparison is finding a forecast technique suitable for routine forecasts of many items. Clearly, a search is inappropriate for each of them. Instead, use will be made of Winters' findings concerning the best composite parameter values obtained during his investigation.

The evaluation method used [Ref. 10] to arrive at the best composite values was to express the forecast error standard deviation for each parameter combination as a percentage above the minimum standard deviation achieved

each product, and to add across the three products for each parameter combination. The best composite rating of 24% was achieved by parameter set (.2, .1) for (α, α'). These parameters will be used in this forecast comparison with the observation that they may not be the best parameters for the particular series used here. As noted above, however, for mass forecasts these parameters must suffice. The robustness or insensitivity of the model to various series should become apparent regardless of the parameters chosen.

Although neither Holt nor Winters gave a procedure for optimal parameter estimation, Harrison [Ref. 7] develops optimal parameters for a specific linear growth model

$$Y_t = \bar{Y}_t + \varepsilon_t \quad .$$

$$\bar{Y}_t = \bar{Y}_{t-1} + b_t + \gamma_t$$

$$b_t = b_{t-1} + \delta_t$$

for which Holt's predictor gives the least mean squared error. Following Harrison's procedure, the difference equation form of the Holt-Winters' Predictor given by

$$\hat{Y}_{t+1} = \hat{\bar{Y}}_t + \hat{b}_t$$

$$\hat{\bar{Y}}_t = \hat{\bar{Y}}_{t-1} + \hat{b}_{t-1} + \alpha e_t$$

$$\hat{b}_t = \hat{b}_{t-1} + \alpha' e_t$$

may be written more compactly as

$$e_{t+1} = ae_t + be_{t-1} - E_{t+1} \tag{D-7}$$

where

$$a = 2 - \alpha - \alpha' \tag{D-8}$$

$$b = -(1-\alpha) = -\beta \tag{D-9}$$

$$E_{t+1} = \varepsilon_{t+1} - 2\varepsilon_t + \varepsilon_{t-1} + \gamma_{t+1} - \gamma_t + \delta_{t+1}. \tag{D-10}$$

Expanding Equation (D-7) gives

$$\varepsilon_{t+1} = \sum_{i=0}^{\infty} W_i \; E_{t+1-i} \tag{D-11}$$

where

$$W_k = a\, W_{k-1} + b\, W_{k-2}$$

and where

$$W_o = 1, \; W_1 = \lambda_1 + \lambda_2 = a, \; W_2 = a^2 + b.$$

When λ_1 and λ_2 are roots of the equation

$$z^2 - az - b = 0$$

it is in general true that

$$W_{k-1} = \frac{\lambda_1{}^k - \lambda_2{}^k}{\lambda_1 - \lambda_2} .$$

From Equation (D-10) and Equation (D-11) it follows that the error variance of forecasts is given by

$$V(e) = \sum_{k=0}^{\infty} [(\Delta_2 W_k)^2 V(\varepsilon) + (\Delta W_k)^2 \; V(\delta) + W_k^2 \; V(\delta)].$$

This variance error tends to a limit if $|W_k| < 1$, which implies that $|\lambda_1|$ and $|\lambda_2|$ are less than unity also. For stability of the expression it follows that

$$\left| \frac{a \pm (a^2 + 4b)^{\frac{1}{2}}}{2} \right| < 1.$$

When a substitution is made for a and b the stability conditons become

$$0 < \alpha' \leq \frac{(\alpha+\alpha')^2}{2} \qquad \text{(D-12)}$$

$$2\alpha + \alpha' < 4. \qquad \text{(D-13)}$$

The optimal parameters must satisfy these conditions, and since they will be determined as functions of the variances $V(\varepsilon)$, $V(\gamma)$ and $V(\delta)$ the parameters are further restricted by

$$0 < V(\varepsilon),\ V(\gamma),\ V(\delta) < V(\min).$$

This implies that the optimal parameters will be contained within the subset of the stability region defined by

$$0 < \alpha' \leq \frac{(\alpha+\alpha')^2}{2+\alpha+\alpha'} < 1 \qquad \text{(D-14)}$$

$$0 < \alpha < 1$$

To determine the optimal parameters, Equation (D-7) can be multiplied by both e_{t+1} and e_t, giving two equations which may be written

$$(1-b^2)S^2 = a(1+b)c_1 + bd_2 + d_o \qquad \text{(D-15)}$$

$$aS^2 = (1-b)c_1 - d_1$$

where $d_j = E[e_{t-j}E_t]$, S^2 and c_1 are the variance and first covariance of the errors, C. Eliminating S^2 from the equation gives

$$(1+b)[(1-b)^2 - a^2]c_1 = (1-b^2)d_1 + abd_2 + ad_o \quad \text{(D-16)}$$

It is shown in Appendix 1 to Ref. 7 that the constants a, b, and V satisfy the stability conditions for the following when defined in such a way that

$$\begin{aligned} bV &= V(\varepsilon) = d_2 \\ aV &= (4-a)V(\varepsilon) + V(\gamma) = -d_1 \\ V &= (6-4a+a^2+b)V(\varepsilon) + (2-a)\,V(\gamma) + V(\delta) = d_o \\ d_j &= 0 \text{ for } |j| > 2. \end{aligned} \quad \text{(D-17)}$$

Given these values, the right hand side of Equation (D-16) simplifies to

$$V[-a(1-b^2)+ab^2 + a] = 0. \quad \text{(D-18)}$$

Referring to Equation (D-8) and Equation (D-9) and substituting for a and b in the above, making use of the stability restrictions in Equation (D-14), it is seen that the coefficient of c_1 in Equation (D-16) is

$$\alpha\alpha'[4-2\alpha-\alpha'] > 0.$$

Since Equation (D-16) equals zero by Equation (D-18), it follows that $c_1 = 0$, and from Equation (D-15) and (D-17) it follows that $S^2=V$. Since all other covariances can be shown to be zero by multiplying Equation (D-7) by e_{t-i}, $i=1,\ldots\infty$, it follows that the predictor is optimal when the parameters satisfy

$$\begin{aligned} V(\varepsilon) &= (1-\alpha)V \\ V(\gamma) &= (\alpha^2+\alpha\alpha'-2\alpha')V \\ V(\delta) &= \alpha'^2 V. \end{aligned}$$

Here V is the minimum variance of forecast error. The problem remains, however, of estimating the variances of the random components of the process. Harrison suggests the use of Serial Variation Curve analysis of the observed data to assess the suitability of the model for forecasting the data and to provide an estimate of the variances needed to select the optimal parameters. By examining the first differences of the data, Harrison [Ref. 7, p. 833] states that the Serial Variation Curve of the first differences is expected to be a straight line with gradient $V(\delta)$ for $i \geq 2$, or

$$E[\Delta Y_t - \Delta Y_{t-1}]^2 = 4V(\varepsilon) + 2V(\gamma) + i\,V(\delta), \quad i \geq 2$$
$$= 6V(\varepsilon) + 2V(\gamma) + V(\delta), \quad i=1.$$

He goes on to state, however, that for practical applications, this method for determining the variances has a large error, and concludes here, as before in his derivation of parameters for the simple EWMA model, that the optimal parameters are again best determined by simulation (empirically) using the actual data. For purposes of this comparison, then, the parameter set recommended by Winters is still retained.

b. The Theil-Wage Model

To determine the adaptation parameters α and α' which minimize the mean square forecast error the forecast error must first be expressed as

$$e_t = \hat{Y}_t + \hat{b}_t - Y_t = \hat{Y}_t + \hat{b}_t - (\bar{Y}_{t-1} + b_{t-1} + \delta_t + \varepsilon_t)$$

where the value of Y_t in Equation (D-1) is substituted into Equation (D-2). This may be written in more concise notation as

$$e_t = A_t + B_t - \varepsilon_t - \delta_t \tag{D-19}$$

where

$$A_t = \hat{Y}_t - \overline{Y}_{t-1}\ , \ B_t = \hat{b}_t - b_{t-1}.$$

To insure that the notation is clear, recall that the forecast of Y_t is $F_t = \hat{Y}_t + \hat{b}_t$ which can be considered as an estimator of Y_t expressed as

$$Y_t = \overline{Y}_{t-1} + b_{t-1} + \varepsilon_t + \delta_t$$

From Equation (D-19) it can be seen that the forecast error is the sum of three terms:

the sampling error (A_t) of the mean process level at time t-1,

the sampling error (B_t) of the slope change at time t-1,

and a disturbance combination associated with the t^{th} period.

Using the forecast model Equation (D-3) and Equation (D-4), the sampling errors can be eliminated successively from Equation (D-19) and it can then be written as

$$\hat{Y}_t - \overline{Y}_t - (\hat{Y}_{t-1} - \overline{Y}_{t-1}) - (b_{t-1} - \overline{Y}_t + \overline{Y}_{t-1}) = -\alpha e_{t-1}.$$

Further rearrangement reveals that A_t can be written as

$$A_t = (1-\alpha)(A_{t-1} + B_{t-1}) + \alpha\varepsilon_t - (1-\alpha)\delta_t. \tag{D-20}$$

The same procedure permits B_t to be written

$$B_t = -\alpha\alpha' A_{t-1} + (1-\alpha\alpha')B_{t-1} + \alpha\alpha'\varepsilon_t - (1-\alpha\alpha')\delta_t. \tag{D-21}$$

Finally these results may be expressed in vector form as

$$\begin{bmatrix} A_t \\ B_t \end{bmatrix} = P \begin{bmatrix} A_{t-1} \\ B_{t-1} \end{bmatrix} + Q \begin{bmatrix} \varepsilon_t \\ \delta_t \end{bmatrix} \qquad \text{(D-22)}$$

where

$$P = \begin{bmatrix} \beta & \beta \\ \theta & \theta' \end{bmatrix}, \quad Q = \begin{bmatrix} \alpha & -\beta \\ \theta & -\theta' \end{bmatrix}, \qquad \begin{matrix} \theta = \alpha\alpha' \\ \theta' = 1 - \theta \end{matrix} \quad .$$

Analysis of the above reflects that an optimal α and θ imply an optimal α' also. By successive elimination of the vector

$$\begin{bmatrix} A_{t-1} \\ B_{t-1} \end{bmatrix}$$

in Equation (D-10) that equation then may be written

$$\begin{bmatrix} A_t \\ B_t \end{bmatrix} = P^n \begin{bmatrix} A_{t-n} \\ B_{t-n} \end{bmatrix} + \sum_{k=0}^{n-1} P^k Q \begin{bmatrix} \varepsilon_{t-k} \\ \delta_{t-k} \end{bmatrix} . \qquad \text{(D-23)}$$

When α and α' are positive (the latent roots of P are within the unit circle), the first term on the right of Equation (D-23) converges to zero when $n \to \infty$. Combining this equation with Equation (D-19) the forecast error may be written

$$e_t = [1 \ 1] \sum_{k=0}^{\infty} P^k Q \begin{bmatrix} \varepsilon_{t-k} \\ \delta_{t-k} \end{bmatrix} - \varepsilon_{t+1} - \delta_{t+1} \qquad \text{(D-24)}$$

On expanding and taking the expectation formally the mean square error is seen to be

$$MSE = [1\ 1](QDQ'+PQDQ'P'+P^2QDQ'P'^2+\cdots)\begin{bmatrix}1\\1\end{bmatrix} + \sigma_\varepsilon^2 + \sigma_\delta^2 \quad \text{(D-25)}$$

where

$$D = \begin{bmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_\delta^2\end{bmatrix}.$$

To simplify the MSE expression, the term in parenthesis may be grouped as $S = QDQ'+PQDQ'P'+P^2QDQ'P'^2+\cdots$ and $PSP' = PQDQ'P' + P^2QDQ'P'+\cdots$ may be subtracted from both sides of Equation (D-25) giving

$$S - PSP' = QDQ'. \quad \text{(D-26)}$$

The result can then be regarded as a set of linear equations in the three elements S_{11}, S_{12}, S_{22} of S since it is seen that S is symmetric. Equation (D-26) can then be written in explicit form as follows

$$\begin{bmatrix}S_{11} & S_{12}\\ S_{12} & S_{22}\end{bmatrix} - \begin{bmatrix}\beta & \beta\\ -\theta & \theta'\end{bmatrix}\begin{bmatrix}S_{11} & S_{12}\\ S_{12} & S_{22}\end{bmatrix}\begin{bmatrix}\beta & -\theta\\ \beta & \theta'\end{bmatrix} =$$

$$\begin{bmatrix}\alpha & -\beta\\ \theta & -\theta'\end{bmatrix}\begin{bmatrix}\sigma_\varepsilon^2 & 0\\ 0 & \sigma_\delta^2\end{bmatrix}\begin{bmatrix}\alpha & \theta\\ -\beta & -\theta'\end{bmatrix}$$

which after rearrangement, has the solution

$$\begin{bmatrix}S_{11}\\ S_{12}\\ S_{13}\end{bmatrix} = \frac{1}{\alpha\theta(4-2\alpha-\theta)}\begin{bmatrix}(2-\alpha\theta)\theta & 2\beta^2\theta & \beta^2(1+\beta)\\ -\beta\theta^2 & (2\alpha(1+\beta)-\theta)\theta & \beta(\alpha(1+\beta)-\theta)\\ (1+\beta)\theta^2 & 2(\alpha^2-2\alpha+\theta)\theta & \alpha^2(1+\beta)+2\theta\beta\end{bmatrix}\cdot$$

$$\begin{bmatrix}\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\delta^2\\ \alpha\theta\sigma_\varepsilon^2 + \beta\theta'\sigma_\delta^2\\ \theta^2\sigma_\varepsilon^2 + \theta'^2\sigma_\delta^2\end{bmatrix}. \quad \text{(D-27)}$$

Returning to Equation (D-26) it is noted that S must be premultiplied by [11] and post multiplied by [11]'. This suggests the form $S_{11} + 2S_{12} + S_{22}$ so Equation (D-27) is premultiplied by [121]. The product of this vector with the 3x3 matrix on the right side of Equation (D-27) results in $[2\theta \;\; 2\theta \;\; 1+\beta]$.

Combining this with Equation (D-26) the mean square prediction error is found to be

$$MSE = S_{11} + 2S_{12} + S_{22} + \sigma_\varepsilon^2 + \sigma_\delta^2 = \sigma_\varepsilon^2 \frac{(4\alpha+2\theta)}{\alpha(2(1+\beta)-\theta)}$$

$$+\sigma_\delta^2 \frac{(1+\beta)}{\alpha\theta(2(1+\beta)-\theta)} = \sigma_\varepsilon^2 \frac{(g^2(1+\beta)+4\alpha\theta+2\theta^2)}{\alpha\theta(2(1+\beta)-\theta)} \qquad \text{(D-28)}$$

where

$$g^2 = \sigma_\delta^2 / \sigma_\varepsilon^2 \qquad \text{(D-29)}$$

the ratio of the slope change variance to the additive error term variance. The nature of the random processes normally encountered suggests that g is often a small number. Now that the MSE has been obtained in fairly simple form, the optimal value of θ may be found by taking the derivative of Equation (D-28) with respect to θ and equating the numerator to zero. Solving the resulting quadratic equation in θ, the optimal θ value may be written

$$\tilde{\theta} = h^2(1+\beta) \qquad \text{(D-30)}$$

where

$$h^2 = -1/8\ g^2 + \frac{1}{2}g(1+1/16g^2)^{\frac{1}{2}} \cdot \qquad \text{(D-31)}$$

Equation (D-30) gives the value of θ which will minimize the mean square forecast error, given a specific α. Recalling

that $\theta=\alpha\alpha'$ it is seen that θ is a decreasing linear function of α. As a matter of convenience it has been found easier to work with h than with g so solving Equation (D-31) for g results in

$$g^2 = 4h^4/(1-h^2). \tag{D-32}$$

Substitution of the θ value of Equation (D-30) into Equation (D-29) gives

$$\underset{\theta}{\text{MIN MSE}} = \sigma_\varepsilon^2 \frac{(g^2+4h^4+2h^2(2-h^2)\alpha)}{h^2(2-h^2)\alpha(1+\beta)}$$

$$= \sigma_\varepsilon^2 \frac{4h^2/(1-h^2)+2\alpha}{\alpha(1+\beta)} . \tag{D-33}$$

Differentiation of Equation (D-33) with respect to α and putting the result equal to zero gives a quadratic equation in α. The results are found to be

$$\alpha = \frac{2h}{1+h} , \beta = \frac{1-h}{1+h} , \quad \alpha' = h \quad \theta= \frac{2h^2}{1+h} \quad \beta' = 1-h \tag{D-34}$$

A substitution of the results of Equation (D-34) into Equation (D-33) gives

$$\underset{\alpha,\theta}{\text{MIN MSE}} = \sigma_\varepsilon^2 \frac{1+h}{1-h} = \sigma_\varepsilon^2/\beta. \tag{D-35}$$

This result has been tabulated in Reference 11 for various values of g^2. For an illustration of the method, however, the optimal parameters for the Theil-Wage forecast model will be computed here, assuming that σ_ε^2 and σ_δ^2 of Equation (D-1) are 25/3 and 4/3 respectively.

$$\varepsilon_t \sim u(-5,5) \Rightarrow \sigma_\varepsilon^2 = (5+5)^2/12 = 25/3$$

$$\delta_t \sim u(-2,2) \Rightarrow \sigma_\delta^2 = (2+2)^2/12 = 4/3$$

from which

$$g^2 = \sigma_\delta^2 \;/\; \sigma_\varepsilon^2 = .16$$

and

$$g = 0.4$$

from Equation (D-30). From Equation (D-13) it follows that

$$h^2 = -1/8\ (16/100)+\tfrac{1}{2}(4/10)(1+.16/16)^{\frac{1}{2}} = 0.181$$

$$h = .425 = \alpha'$$

$$\alpha = 2(.425)/1.425 = .850/1.425 = 0.596.$$

These values of α and α' are those used in the computer comparison of the Theil-Wage model. These parameters will insure that the model is optimal for the series generated in accordance with Equation (D-1) where the range on ε_t is 10 and the range on δ_t is 4. Using the results of Equation (D-35) it is noted that the min. M.S.E. should be approximately 20.6.

E. BROWN'S ONE-PARAMETER EWMA LINEAR GROWTH MODEL

1. The Series Generation Model

The series generation model for Brown's forecast rule is simply the forecast rule with random numbers applied as the shocks to process level and slope. Comparison with Equation (E-5) will reveal that the forms are equivalent. The model may be written

$$\begin{aligned} Y_t &= \bar{Y}_t + \varepsilon_t \\ \bar{Y}_t &= \bar{Y}_{t-1} + b_t + \varepsilon'_t \qquad \text{(E-1)} \\ b_t &= b_{t-1} + \varepsilon''_t \end{aligned}$$

where Y_t is the observation at time t, $\overline{Y}_t$ is the process level at time t, b_t is the process slope at time t and ε_t is the random shock observed as a forecast error at time t. ε'_t is a random shock experienced by the process level, and Brown's forecast rule writes this shock in terms of the forecast error $\varepsilon'_t = (1-\beta^2)\varepsilon_t$. ε''_t is the random change in process slope which Brown's model assumes may be written $\varepsilon''_t = \alpha^2\varepsilon_t$. This model is essentially the linear growth model in Equation (D-1) with Brown's assumptions concerning the relationship of the various random elements.

2. <u>The Forecast Model</u>

Brown's forecast model as discussed in Reference 7, is essentially a special case of the Holt-Winters model Equation (D-3) and Equation (D-4) in which he restricted the second parameter to be a function of the first. The forecast rule is identical to Equation (D-2)

$$F_t = \hat{Y}_t + k\,\hat{b}_t \cdot \tag{E-2}$$

To illustrate the similarity of Brown's model to the Holt-Winters model, the latter model may be written in the random shock form using the forecast error as the shock

$$\varepsilon_t = Y_t - (\hat{Y}_{t-1} + \hat{b}_{t-1}).$$

Equation (D-3) becomes

$$\hat{Y}_t = \hat{Y}_{t-1} + \hat{b}_{t-1} + \alpha\varepsilon_t \tag{E-3}$$

and Equation (D-4) becomes

$$\hat{b}_t = \hat{b}_{t-1} + \alpha\alpha'\varepsilon_t \tag{E-4}$$

When Brown's double smoothing model is written in the same form [Ref. 7], the similarities are easily seen as

$$\hat{Y}_t = \hat{Y}_{t-1} + \hat{b}_{t-1} + (1-\beta^2)\varepsilon_t \qquad \text{(E-5)}$$

$$\hat{b}_t = \hat{b}_{t-1} + (1-\beta)^2 \varepsilon_t .$$

The Holt-Winters parameter α, which smooths the process level in Equation (E-3) is equivalent to Brown's parameter $(1-\beta^2)$ or $\alpha(2-\alpha)$. The Holt-Winters' parameter product α' in Equation (E-4) is equivalent to Brown's $\alpha/2-\alpha$. When these equivalent parameters are numerically equal, then Brown's model will yield the same result as the Holt-Winters model. The forecast model used in the computer program for this comparison, however, is the more familar from of Brown's double smoothing, namely,

$$\hat{Y}_t = 2\tilde{Y}_t - \tilde{Y}_t^{(2)} \qquad \text{(E-6)}$$

$$\hat{b}_t = \frac{\alpha}{\beta} [\tilde{Y}_t - \tilde{Y}_t^{(2)}] \qquad \text{(E-7)}$$

where

$$\tilde{Y}_t = \alpha Y_t + \beta \tilde{Y}_{t-1} \qquad \text{(E-8)}$$

$$\tilde{Y}_t^{(2)} = \alpha\tilde{Y}_t + \beta\tilde{Y}_{t-1}^{(2)} . \qquad \text{(E-9)}$$

Equation (E-6) expresses the estimate of the lag-corrected current process level, and Equation (E-7) expresses the current estimate of the process slope. Equations (E-8) and (E-9) are the single and double smoothed expressions used to calculate the level and slope. The method involved in these calculations makes use of the known fact that single smoothing (the simple exponentially weighted moving average of Equation

(E-8)) lags a trend by a constant amount. The double smoothed version in Equation (E-9) lags the single smoothed value in the same amount by which single smoothing lags the observed trend. If Equation (E-6) is rewritten as $\hat{Y}_t = \tilde{Y}_t + (\tilde{Y}_t - \tilde{Y}_t^{(2)})$ the term in parenthesis is recognized to be the lag correction required to give the current adjusted estimate of process level. The amount of lag inherent in Equation (E-8) for trend series is the observed result of using past data to estimate the current level. This lags corresponds to the "average age" of the past data used in making the estimate, and the error due to this lag can be shown to be

$$e_t = \beta/\alpha \; b_t$$

therefore the estimate of slope can be written as Equation (E-7), where $e_t = \tilde{Y}_t - \tilde{Y}_t^{(2)}$. Brown extends this approach further to form what he terms third order smoothing, and higher levels, but these forms are not relevant to the present comparison.

3. Parameter Estimation

Brown advocates an empirical approach in "fitting" the forecast to the series, but he also says that α=0.1 is a good multi-purpose value. Since the Holt-Winters parameter values were taken from the literature as their best estimate of a good value combination, Brown's value will be taken as such also. While this value may be improved upon for specific series by experimentation, this superficial selection is in keeping with the objective of this study - to forecast varied types of series at minimum cost.

F. THE BOX-JENKINS POLYNOMIAL PREDICTOR - THE GENERAL MODEL OF ORDER N

1. The General Polynomial Growth Model

The generating model for which the Box-Jenkins Polynomial predictor is the optimal linear least square predictor is discussed by Harrison in Reference 7. The general polynomial model specifies that each "derivative" of the underlying process experiences a random change in such a way that for the N order polynomial the model is given by:

$$Y_t = \overline{Y}_t^{(1)} + \varepsilon_t \tag{F-1}$$

$$\overline{Y}_t^{(i)} = \overline{Y}_{t-1}^{(i)} + \overline{Y}_t^{(i+1)} + \gamma_t^{(i)}, \quad i=1,\ldots,n. \tag{F-2}$$

Here $\overline{Y}_t^{(n+1)} = 0$ and the random errors ε_t and $\gamma_t^{(i)}$ have mean zero and variance $\sigma^2 I$. Comparison of Equation (F-1) and Equation (F-2) with the steady model, Equation (B-4), reveals that the latter equation results from i=1 in Equations (F-1) and (F-2). A similar comparison of the linear growth model Equation (D-1), reflects that this form results from setting i=1,2 in Equation (F-2). The particular model in Equation (D-1), has $\overline{Y}_t^{(2)} = b_t$ and $\gamma_t^{(1)} = 0$ but $\gamma_t^{(1)}$ could easily have had a positive value and nothing in that analysis would have changed except the complexity of terms in the development.

2. Optimal Properties of the Polynomial Predictor

For the assumed generating model the Box-Jenkins polynomial predictor can be shown to be optimal in the linear least squares sense and the optimal values of the

forecasting parameters α_i can be derived as functions of the variances of the random shocks ε_t and $\gamma_t^{(i)}$. A special case of the Box-Jenkins result derived on pp. 312, 313 of Reference 14 will be employed and the procedure followed by Harrison on p. 824 of Reference 7 will be used to show the result.

Let a stochastic process by generated by the model

$$Y_{t+1} = \sum_{j=0}^{\infty} \eta_j Y_{t-j} + \varepsilon_{t+1} \tag{F-3}$$

where the $\{\varepsilon_t\}$ have zero mean and are identically distributed. uncorrelated random variables, and the η_j are constants. Consider a forecast rule of the form

$$\hat{Y}_{t+1} = \sum_{j=0}^{\infty} \mu_j Y_{t-j} \tag{F-4}$$

where the μ_j are constants. Since it is true that

$$E[e_{t+1}]^2 = E[Y_{t+1}-\hat{Y}_{t+1}]^2 = E[\sum_{j=0}^{\infty} (\eta_j-\mu_j)Y_{t-j}]^2+E[e_{t+1}]^2 \tag{F-5}$$

the forecast rule will be optimal in the least squares sense when $\mu_j = \eta_j$. When this is true, Equation (F-5) implies that the forecast errors $\{e_t\}$ are identical to the random shocks $\{\varepsilon_t\}$, and therefore are also uncorrelated. By successive substitution, the Equation (F-4) form of the forecast rule may be transformed in terms of the forecast errors as

$$\hat{Y}_{t+1} = \hat{Y}_t + \sum_{j=0}^{\infty} W_j e_{t-j} \tag{F-6}$$

where the W_j are constants. Writing Equation (F-6) in difference equation form

$$\nabla\hat{Y}_{t+1} = \sum_{j=0}^{\infty} W_j \, e_{t-j} \quad \text{(F-7)}$$

it follows from the previous analysis that this forecast rule is optimal for the equivalent underlying stochastic process

$$\nabla Y_{t+1} = \sum_{j=0}^{\infty} W_j \, \varepsilon_{t-j} + \nabla\varepsilon_{t+1}$$

since the forecast errors were equivalent to the random shocks. Specifically, the forecast rule is optimal for a series generated by the model

$$\nabla Y_{t+1} = \sum_{j=0}^{n-1} \gamma_j \, S^j \, \varepsilon_t + \nabla\varepsilon_{t+1} \quad \text{(F-8)}$$

where S^j is used to denote the j^{th} multiple sum. Differencing Equation (F-8) n-1 times, Harrison arrives at the expression

$$\nabla_n Y_{t+1} = \sum_{j=0}^{n-1} \gamma_j \, \nabla_{n-j-1} \, \varepsilon_t \quad \text{(F-9)}$$

which has no error terms beyond ε_{t-n+1}. The result which has been found states that if a random observation y has the property that its n^{th} difference can be represented as a moving average process of i.i.d. variables $\{\varepsilon_t\}$ which have zero means and the process is of order n+1, then the Box-Jenkins n order polynomial predictor given by

$$\hat{Y}_{t+1} = \hat{Y}_t + \sum_{i=0}^{n-1} \eta_i \, S^i \, e_t \quad \text{(F-10)}$$

is optimal and the $\{\varepsilon_t\}$ represent the one-step-ahead forecast errors for this optimal rule.

3. Parameter Estimation

The procedure used by Harrison [Ref. 7] shown in section B-5 for the simple EWMA model demonstrates the Box-Jenkins parameter estimation procedure for the steady model. The optimal predictor for the linear growth model is the form of the general predictor recommended by Holt-Winters. Harrison has derived [Ref. 7] the optimal parameters for this model as previously described in Section D.3.a.

4. The Relationship of Special Cases of the General Polynomial Predictor

The forecast models proposed by Holt-Winters, Theil-Wage, and Brown are specific forms of the Box-Jenkins Polynomial Predictor. The demonstrate the relationship, Equation (F-10) can be written in the form

$$\hat{Y}_{t+1} = \sum_{i=1}^{n} \hat{Y}_t^{(i)} \tag{F-11}$$

where

$$\hat{Y}_t^{(i)} = \sum_{j=1}^{n} \hat{Y}_{t-1}^{(j)} + \alpha_i e_t \tag{F-12}$$

and

$$e_t = Y_t - \hat{Y}_t$$

This form resulted from expressing the $\hat{Y}_t$ in terms of the errors and substituting in Equation (F-11).

a. The Simple Exponentially Weighted Moving Average

The first order predictor, where n=1 in Equation (F-11) causes that expression to reduce to the simple EWMA model

$$\hat{Y}_{t+1} = \hat{Y}_t$$

$$\hat{\hat{Y}}_t = \hat{\hat{Y}}_{t-1} + \alpha e_t$$

where equivalence to Equation (B-3) is seen by substitution for the error in terms of the observation and forecast,

$$\hat{\hat{Y}}_t = \hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\, \hat{Y}_t .$$

b. The Holt-Winters Linear Growth Model

When n=2 in Equation (F-11), the second order predictor becomes

$$\hat{Y}_{t+1} = \hat{\hat{Y}}_t + \hat{b}_t$$

$$\hat{\hat{Y}}_t = \hat{\hat{Y}}_{t-1} + \hat{b}_{t-1} + \alpha\, e_t$$

$$\hat{b}_t = \hat{b}_{t-1} + \alpha' e_t$$

where $\hat{Y}_T^{(2)}$ is written as $\hat{b}_T$.

c. The Brown Linear Growth Model

Brown's second order predictor is a particular form of Holt's, where the two parameters α, α' are restricted to be functions of the discounting parameter β,

$$\alpha = 1-\beta^2$$

$$\alpha' = (1-\beta)^2 .$$

d. The Theil-Wage Linear Growth Model

The Theil-Wage Predictor restricts the Holt-Winters Parameters so that

$$\alpha = \frac{2h}{(1-h)}, \quad \alpha' = h\alpha .$$

III. RELATIVE FORECAST MODEL PERFORMANCE COMPARISON

A. COMPARISON METHODOLOGY

1. Time Series Generating Model Specification

When forecasts of a particular time series are attempted, the forecaster's belief concerning the underlying generating process governs his selection of a specific forecast model. Although recognizing that few models completely describe the complexity of practical economic or physical processes, by his selection of a forecast model the forecaster thereby indicates that the time series has an underlying stochastic process whose functional form is suggested by the forecast rule. The adaptive parameters chosen for use tend to further specify the assumed form of the generating process. The degree to which the actual underlying process differs from the assumed form is reflected in the forecast accuracy obtained by the forecast model.

2. Forecast Model Specification Error Measurement

Any comparison of relative forecast accuracy is fundamentally an attempt to determine which of the forecast models is the more accurate specification of the process generating the series being forecast. The measure of specification error used in this comparison will employ the method discussed by Bossons [Ref. 13], who defined specification error as the additional variance of forecast errors introduced by misspecification of the generating process. This measure cannot be determined when comparison is attempted using actual

time series data, since the precise underlying model is unknown. Since the various series examined in this comparison have been generated from known process models for which the optimal forecast model and parameters are also known, it is possible to determine

$$f = \frac{VAR(Y_t - Y_t^{**})}{VAR(Y_t - Y_t^{*})} - 1$$

as a measure of the model specification. If $\{Y_t\}$ is the series being forecast, then $\{Y_t^{**}\}$ is the series of forecasts generated by the forecast model whose associated specification error is being measured, and $\{Y_t^{*}\}$ is a series of optimal (minimum asymptotic variance) forecasts for the series $\{Y_t\}$. The lower bound on f is zero, and is obtained when $\{Y_t^{**}\}$ = $\{Y_t^{*}\}$. No upper bound on f exists. Positive f values reflect the extent to which a particular forecast series has been degraded by misspecification of the underlying stochastic process. This measure, as Bossons [Ref. 13] has observed, has two uses. First, it permits the effect of a known misspecification, such as a model simplification, to be measured. This emphasizes the relative importance of various coefficients or parameters in the model, and its sensitiveness to change. Second, it permits the robustness of a forecast model to be measured by reflecting the forecast accuracy for various types of series misspecifications, such as correlation, linearity, or for specific distributions of the random variables in the process.

3. Forecast Model Performance Criterion

The measure of model specification error permits a preference ordering or ranking of forecast models to be made for any given time series generation model. It is often true that a collection of these series must be forecast, and few of the series are represented by any one generating model. It would, of course, be possible to analyze all the series and group them according to these process relationships, and for some processes this expense is justified. For many series however, as mentioned in the introduction to this study, it is not critical that the forecasts be exceedingly accurate, and the forecast costs must be kept to a minimum, consistent with the benefits obtained from such forecasts. An additional means of measurement, therefore, is needed to permit a preferred ranking of forecast models over a collection of series to be forecast. It may be true that one or two forecast models may provide the necessary accuracy over the entire collection of series. However, it is recognized that this is extremely situation dependent. The method used in this study to suggest possible "best" forecast models for use in predicting a collection of dissimilar time series is the calculation of the average specification error over all series forecast, and the sample variance of the specification error. Selection of a model based on the estimated mean specification error would, of course, imply that a few forecasts which were grossly erroneous could be tolerated, while selection based upon estimated specification error variance would suggest that a

uniformly high level of accuracy was not required but that no extremely poor forecasts were acceptable. Some combination of the two criteria may even be considered. The conclusions of this study will be restricted to rankings in terms of the mean and variance of specification error. No valid general interpretation can be given to these conclusions, since they are applicable only for the class of models and distributions chosen for use in this study.

The method would, however, be applicable to the relative comparison of a group of forecast models when forecasting actual time series, if one were willing to assume that the particular model with the least mean square forecast error was "optimal" for a particular series. Even with specification error based only upon the "best" model, rather than upon a truly optimal (minimum variance error) model, it would still be possible to draw meaningful conclusions concerning the relative effectiveness of a group of models. The same approach would apply when searching for a general purpose parameter for use in a single model which must be applied to a collection of series. It is unlikely that an optimal parameter for one series will be optimal for all. Use of this method on a representative sample of series would facilitate selection of the parameter which minimizes the average forecast error variance or some other selected criterion for the entire collection of series, and not just for one particular series.

4. <u>Time Series Data Generation Methods</u>

a. Parameter Selection

For each of the forecast models discussed in Section II a time series was generated using the underlying process generation model associated with that specific forecast rule. The random shock forms of those rules only require that the parameters, such as intercept and slope, be specified and that the random shock be added to provide the stochastic element for the series. It was determined during the course of the study that the variance of the random element caused little or no change in the specification error, so no attempt was made to produce results over a range of model parameters. The reason for this was that, even though forecast error variance was changed for all models, it was changed in generally the same proportion for all and the ratios remained approximately constant. In no case would the ranking have changed due to the parameters selected. Parameters used to obtain the representative results contained in Tables I and II may be determined by consulting the computer program.

b. The Random Number Generator

The random number generator upon which all stochastic series properties are based is a function called URN which is contained in the Naval Postgraduate School IBM 360-67 computer. This additive generator was selected as a standard instead of available multiplicative generators or the conventional RANDU since it was almost three times faster than RANDU and had been subjected to statistical tests which

were on file at the computer facility at USNPGS. As a known quantity, questions involving this generator and its effect on results should be quickly resolved without additional statistical testing of the generator.

c. Normal Random Number Generation

The normal random numbers used with the Least-Squares Models were generated from the uniform (0,1) URN output by using the Central Limit approach outlined by Naylor [Ref. 8] on pages 92-93. This amounted to summing twelve uniform (0,1) numbers and subtracting six to produce a normal (0,1) number.

d. Uniform Random Number Range Transformation

The uniform random numbers were transformed to various ranges (page 79, Ref. 8) using the expression

$$Y_t = A + (B-A)\varepsilon_t \quad 0 \leq \varepsilon \leq 1$$

which is a rearranged form of the uniform cumulative distribution function. Since A = -B in this study, this expression can be written as

$$Y_t = (2\varepsilon_t - 1)B = (\varepsilon_t - 0.5)2B$$

which is the form used in the Fortran program to transform the random numbers to desired ranges.

5. <u>Single and Group Series Forecast Performance Comparison</u>

a. Single Series Comparison

After a time series was produced from one of the generating models in the computer program, all forecast models

were used to forecast each observed series value, given the past values. After each forecast the forecast error was stored and after all models had completed forecasting the series, the average forecast error and the sample variance of this error were calculated. From this the sample estimate of the specification error (f) discussed in Section 2 above was computed for each model. Table I contains data for representative comparison made using this method of measurement.

b. Forecast Comparisons Using Several Series

When the single series comparisons had been completed, the average specification error and estimated sample variance were then calculated for each forecast model. This is the performance criterion discussed in Section 3. As noted there, the results may be interpreted only in relation to the specific series combination examined. Table III contains representative data for comparison of series using this criterion.

6. <u>Forecast Model Stabilization and Operation</u>

A stabilization period of 100 observations was used to remove any effects due to improper starting conditions and then the forecast error was compiled over the next 300 observations. The specific initial conditions for each model which were used to obtain the representative results in Tables I and II may be determined by consulting the computer

program. Sufficient comments have been provided to assist in identification of program components and the variables used in each. Although the results represent only sample estimates of average forecast error and forecast error variance, the number of observations forecast was considered large enough to provide a reasonable estimate of the actual values.

B. COMPUTER COMPARISON ANALYSIS

1. Prior Performance Expectations

The forecast models compared were selected with certain *a priori* relative outcomes in mind. Due to the nature of the assumptions on the models and the generated series associated with those models it was anticipated that

a. each model would be "optimal" (have the least estimated mean squared error) for its corresponding time series.

b. the modified least-squares forecast model would produce smaller forecast errors than the simple least-squares model due to the correlated nature of most series used in the study.

c. the forecasts generated by the simple EWMA model would tend to lag behind the linear growth processes by a relatively constant amount.

d. the Cox-modified EWMA would tend to track the linear models better than the unmodified model.

e. the Holt-Winters model would probably perform better overall than any of the models above since it is

suitable for both the autocorrelated data and for the linear model used in five of the generating processes.

f. the Theil-Wage model would have the same general characteristics as the Holt-Winters model, but probably provide better response to process changes due to their use of a more "simultaneous" approach (use of the most recent slope rather than the use of the slope calculated last period) than that used by Holt and Winters.

g. the Brown model would have generally the same characteristics as the Holt-Winters model, but not perform quite as well due to the restrictions which Brown placed on forecast parameters. It is argued by some [Ref. 7] that Brown's more parsimonious model can achieve essentially the same results in practice as that produced by either the Holt-Winters model or the Theil-Wage model. Quantitative evidence to support or refute this claim was one expected result of this investigation.

2. Forecast Model Performance Comparisons

a. The Least-Squares Series

The least-squares forecast model performed better than all others in predicting this series, as anticipated. Comparison of the process variance of 9.13 (see Table II) with the forecast error variance of 8.98 tends to confirm that the least-squares forecast model produced optimal predictions. The next best forecast model for this series was Brown's double smoothing model with an 11.4% larger forecast error variance.

It may be observed (Table I) that the modified least-squares model was ranked fourth after the Holt-Winters model. The unnecessary correction for correlation in the random shocks degraded the quality of forecasts generated by this model. Note also that the simple EWMA forecast model lags the observations by 3.036 (see Table I). This was expected since this model specifies that the forecast for the next period (or any future period) is the current level of the process, and slope of the linear model was selected to be 3.0. The EWMA model was thus never able to anticipate the change due to slope. This is the refinement incorporated into Brown's double smoothing method and it appears to be effective in that model.

b. The Modified Least-Squares Series

The results for this series were much the same as for the simple least-squares series except that the least-squares models reversed their roles. The modified least-squares model produced approximately the same forecast error variance (8.98) from the correlated series as the least-squares model had obtained previously on the uncorrelated series (see Table II), and the least-squares model produced almost the same specification error for this series as the modified model had obtained when forecasting the uncorrelated series. The change from uncorrelated to correlated normal random shocks affected the other models in varying degrees (see Table I). The least-squares model and the Brown model appear to possess equivalent capabilities to forecast the

correlated series, while the modified EWMA model continued to be the least desirable for use in forecasting the series.

c. The Simple EWMA Series

The simple EWMA model, although its forecast error variance was not as low as the series variance (see Table II), gave the least forecast error variance for this series. Almost all models seemed to be capable of forecasting this series adequately, but the Theil-Wage model was the least desirable by far. The non-linearity of this series became noticeable in the relatively poor performance of the least-squares model, but the effect was not pronounced due to the limited range of the random walk in the series generating model. The modified least-squares model was more capable of forecasting this series due to its ability to use the added information contained in the series correlation.

d. The Modified EWMA Series

The modified EWMA forecast model developed by Cox generally provided the least forecast error variance, but it may be observed (Table I) that the modified least-squares model produced almost the same results for this series. The modified EWMA model does not share this versatility when forecasting the modified least-squares series. The next ranked model in terms of least forecast error was the simple EWMA model, but its error variance was 18.4% greater than the modified EWMA model. The modified least-squares model provided a 0.006 specification error during

this particular comparison, but on others it reflected a slight negative error implying that it performed better than the "optimal" model. Based on the simulation results there appears to be no real difference in forecast accuracy regardless of which of these two models were used. Reference to Table II reflects that these two models were able to reduce forecast error variance substantially below that observed in the series. This suggests a true predictive capability not possessed by the other models whose forecast error variances were each approximately equal to or greater than the series variance.

e. The Holt-Winters Linear Growth Series

The Holt-Winters forecast model achieved the least forecast error variance as expected, but this performance was only slightly better than that shown by the Brown model (see Table I). A few comparison trials have resulted in the Holt-Winters model obtaining a smaller forecast error variance for the Brown growth series, and the Brown forecast model demonstrates a similar capability to perform better than the Holt-Winters model on the Holt-Winters series. These outcomes were regarded as sample variations, but the implication is obvious. The results of these models are so comparable that it would be difficult to conclude that any real difference existed between them (for this series). The next best performance (the Theil-Wage model) resulted in a 72.6% increase in forecast error variance.

f. The Theil-Wage Series

The Theil-Wage forecast model proved to be distinctly optimal for forecasting its assumed underlying process. Previous comparisons have shown that the various observed series could have been generated by several slightly different models, as evidenced by the comparable forecast performance of the several forecast models. Here, however, the next best forecasts were obtained by the simple EWMA model with almost four times the forecast error variance. The minimum error variance of 18.14 (see Table II) exhibited by the Theil-Wage forecast model is much larger than the series variance of 7.99, but it compares favorably with the predicted theoretical mean square error of 19.7 obtained by use of Equation (D-34).

g. The Brown Linear Growth Series

The results obtained while forecasting this series tend to further reinforce the observations made concerning the Holt-Winters results. The Brown double smoothing model obtained the least forecast error variance, but the Holt-Winters model only slightly exceeded that minimum value. Other models tended to forecast this series somewhat more accurately than the Holt-Winters Series, but the same performance ranking resulted (see Table I).

TABLE I. MODEL SPECIFICATION ERRORS[1] AND AVERAGE FORECAST ERRORS[2]

Forecast Model / Time Series	Least Squares	Modified Least Squares	Simple EWMA	Modified EWMA (Cox)	Holt-Winters	Theil-Wage	Brown
Least Squares	0.0 -0.020	0.174 -0.019	2.023 -3.036	3.825 -5.041	0.147 0.012	1.131 -0.021	0.114 0.007
Modified Least Squares	0.183 -0.019	0.0 -0.017	1.434 -3.030	3.210 -5.014	0.227 0.035	0.690 -0.020	0.180 0.024
Simple EWMA	0.370 0.327	0.074 0.205	0.0 0.071	0.028 0.286	0.042 0.073	0.613 0.012	0.012 0.097
Modified EWMA (Cox)	0.217 0.085	0.006 0.050	0.184 0.015	0.0 0.059	0.305 0.024	0.726 -0.009	0.249 0.025
Holt-Winters	35.270 -16.023	12.898 - 9.68	0.946 -1.667	1.463 -2.696	0.0 -0.197	0.722 -0.024	0.004 -0.353
Theil-Wage	6501.890 172.919	2346.421 103.268	2.723 2.030	2.814 2.170	11.837 0.197	0.0 -0.014	23.061 0.288
Brown	9.489 -5.827	3.442 -3.505	0.754 -1.051	0.962 -1.739	0.009 -0.010	0.692 0.011	0.0 0.006

[1]Model Specification Errors are the Upper Values

[2]Average Forecast Errors are the Lower Values

TABLE II. SUPPLEMENTARY FORECAST MODEL PERFORMANCE DATA

Forecast Model / Time Series	Least Squares	Modified Least Squares	Simple EWMA	Modified EWMA (Cox)	Holt-Winters	Theil-Wage	Brown
Optimal Model Forecast Error Variance *	8.98	8.98	17.84	2.98	8.72	18.14	9.14
Series Variance	9.13	9.10	15.32	3.67	8.44	7.99	8.78
Sample Mean and Variance of Optimal Model Specification Error	935.345	33.573	1.153	1.757	1.795	0.653	3.374
			0.827	2.043	16.818	0.096	64.601

* If forecast error variance for any other forecast model (i) is of interest for some series, it may be determined by the calculation

$$\sigma_i^2 = (f_i + 1)\ \sigma_{opt}^2$$

which is a form of the Equation [(A-1) Section III] used to determine the specification error.

TABLE III. SAMPLE AVERAGES AND VARIANCES OF SPECIFICATION ERROR

Forecast Model	Over all Seven Series Average	Over all Seven Series Variance	Over Six Series Excluding Theil-Wage Average	Over Six Series Excluding Theil-Wage Variance
Least Squares	935.345	xxxxxxxx	7.5882	164.791
Modified Least Squares	337.537	xxxxxxxx	2.7656	22.058
Simple EWMA	1.153	0.827	0.8903	0.482
Modified EWMA	1.757	2.043	1.5812	2.166
Holt-Winters	1.795	16.818	0.1216	0.013
Theil-Wage	0.653	0.0956	0.7622	0.029
Brown	3.374	64.601	0.0931	0.009

IV. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSIONS

1. Validity of Assumptions

Throughout this thesis it has been stressed that the final conclusions would necessarily be general in nature. The assumptions which led to these conclusions must not be overlooked, for some of them may place significant restrictions on the applicability of the methods studied. If these assumptions are changed, introducing other distributions or parameters, the methodology introduced here is equally valid in those circumstances.

Some specific assumption which are thought to affect the conclusions in some manner are:

a. The composition of the selected group of forecast models. Some models not considered here may have proved superior to all of those selected. The conclusions therefore are applicable only to the specific set of forecast models treated.

b. The specific parameters selected for each forecast model. Winters and Brown have recommended certain values for use with their models for general purpose application. No such recommendation was found for the Theil-Wage model, for example. A specific series was generated, and the optimal parameters for that series were used for all series. Perhaps a more general set of parameters for the Theil-Wage model exists which would have made its performance superior to the other models.

c. The use of the uniform distribution and the range selected for each may have had significant impact on forecast model performance. These distributions do not necessarily represent any physical or economic process.

d. Finally, it must be remembered that these conclusions result from simulated time series data where the generating parameters were accurately known, and no sweeping claims can be made for the results of any simulation.

While it is perhaps proper to be skeptical of the specific conclusions to be made here, it is again suggested that much better assumptions can be made in the context of a problem, and when this comparison is repeated, the conclusions at that time should be more significant and of practical value to the forecaster.

2. Conclusions and Applications

Based upon the representative results in Tables I and II, and the analysis in Section III, B, the following conclusions are made concerning forecast model performance for the tested series:

a. The use of the least-squares model and the modified least-squares model for forecasting the Holt-Winters, Theil-Wage, and Brown series gave poor results, primarily because these series tend to be non-linear for many of the chosen ranges of the distribution of the random variables. In attempting a "non-discounted" linear fit to data which appear quadratic over large periods, enormous errors result.

The lesson here, if one may be found, is that a linear forecast model such as least squares must not be applied to data which have a tendency to be non-linear. When such a case is suspected, the "discounted" linear models are to be preferred, since they tend to fit the data only locally, with the more distant observations given relatively little weight.

(b) The performance of the Holt-Winters, Brown, and modified least-squares linear models on stationary (no trend) data is good, so unless no question exists as to the constant level of the random process it would appear desirable to use a linear model. In the event that a trend occurs, the model will follow it well, and if not, it will still give good forecasts. If a series similar to the modified EWMA is likely to be encountered, the modified least-square is to be preferred over the Holt-Winters or Brown models. This model is distinctly superior to all others when only the first four series in Table I are considered. One reason for its superiority over the standard least-squares model is that the latter can generally only be expected to have, at best, a forecast error variance equal to the variance of the process about its mean. This is due to the model's attempt to forecast the mean value of the series (the least-squares regression line). The modified form further extends the simple least-squares model and adds a correction factor based on the known correlation and the last forecast error. With this added information its forecast capability is much improved. That

this is only true when the random shocks are correlated, however, may be seen from Table I.

c. The Modified EWMA Model was designed to permit Wiener's linear predictor to adapt to changes in process level. It did not appear to achieve this goal very effectively, since the simple EWMA model performed better on every series except the one for which the modified model was intended to be optimal. The data tend to suggest that the Cox modified model is a special one with limited adaptive capability when applied as a result of misspecification of the underlying model. This, of course, is not in agreement with the prior expectations of this model as stated in Section IV B. 1, but after some reflection is not very surprising. The model was intended to reduce bias caused by a change in the mean level of a random walk. The rapid changes which occur in the linear model would not normally be expected to occur in a random walk process, or the walk would lose its random property. Therefore the model was not intended to follow a linear trend, and without further modification should not be used to attempt this.

d. The results in Table I lead to a conclusion that the Brown and Holt-Winters models are comparable, with Brown's model showing slightly better overall performance. This tends to support Harrison's claim [Ref. 7] that Brown's model is preferred in practice due to its simpler construction but comparable result.

e. The measure suggested earlier in this study of an average specification error, and specification error

variance in selecting the best models tends to bias selection in favor of the Theil-Wage forecast model (see Table III). The reason for this is that, although it did not perform as well on all other series as for instance the Brown model, the other models almost completely failed to follow the Theil-Wage series. As a result, the variance and average of specification error were inflated for all other models. The uniqueness of the Theil-Wage series, as evidenced by the relatively poor showing of other forecast models in predicting this series, is an excellent example of the need to make such model comparisons as this study has done. If one were restricted to using the other forecast models for predicting this series, it might be thought that the series was too violent in its fluctuations to be accurately forecast by any model. When such a series occurs in practice and none of the standard models seem to apply, a comparison of widely assorted forecast models may suggest a more appropriate form for use.

An application of the comparison methodology presented in this thesis might be made by a supply item manager, who is responsible for forecasting demands for stocked items and insuring that stockouts do not occur more frequently than some specified rate. The manager could select sample demand data which was representative of his stockage items, or if the differences in demand distributions for some items were too great, he might form two or more homogeneous groupings from which representative series were taken and select a

model for each. The forecast models could then be used to forecast the selected series and the measures of specification error calculated using the least forecast error variance model which resulted. Of course, the forecast technique in use at the time of the comparison should also be included. If the method is a good one, this comparison will demonstrate clearly how good it is in terms of the added forecast error involved in using other less applicable methods. The interpretation of 100 times the specification error as the added percentage of forecast error variance caused by model misspecification should be easily understood by all those involved in model selection. This approach is recognized to be strictly empirical, but for mass producing forecasts on a routine basis at least cost, further analytical study may not be justified in many cases.

In the event that no model appears clearly superior to others, the comparison results still permit an intelligent model selection procedure. Consider the results in Table I. It is evident that when the linearity, serial independence and normality assumptions are satisfied the least squares method should always be used. Uncertainty about the satisfaction of these assumptions poses an interesting decision problem. If the least squares forecast model is used and the generating model of the forecast series is actually associated with one of the other forecast models, the forecast error variance "penalty" for use of the least squares method ranges from an additional 18.3% minimum to over 650,000%.

On the other hand, if one were to select the Brown model, the "penalty" would range from 11.4% if the least squares model should have been chosen, up to a maximum of over 2300% in a worst case. For six of the seven possible series, however, the Brown model would result in less than a 25% penalty. The "max-min" solution suggested by Table I is the selection of the Theil-Wage model, which never results in a forecast error variance penalty greater than 113% or just over twice the error variance, no matter which series is forecast. It may be noted that this model, though, always causes at least a 60% penalty unless it happens to be the optimal model. If the forecaster had reasonably good information on the likelihood of occurrence of the various series and some loss function in terms of forecast error variance, a decision rule could be formulated to guide the selection of a preferred forecast model.

An additional application which is actually only a variation of the approach used in this thesis, is the modification of various generating models to introduce varying degrees of autocorrelation (or lagged variables in the least squares models) to determine their relative effects on the forecast models. Previous studies have shown a negative bias in variance of least squares parameter estimates caused by autocorrelation. A negative bias in the estimates themselves occurs when $\beta > 0$ in the lagged variable case. These results may be shown analytically (p 211-221, [Ref. 12]). The studies have gone on to show, through simulation, that

a strong positive bias occurred when autocorrelation and lagged variables were both present. Such investigations of model interactions could make use of the methodology described in this thesis as a quantitative measure of results, and further add to the present understanding of economic time series data.

B. RECOMMENDATIONS

1. Extensions of this Investigation

Some recommended extensions of this thesis which may result in improved forecasting models or a better forecast model selection criteria are:

a. A determination of the effect on the results of this study of distributions other than uniform.

b. Investigation, along the lines suggested by Cox [Ref. 5], of the advantages of replacing $\hat{Y}_t$ in Equation (C-3) by some other form of moving average.

It would be interesting to determine whether requirements such as minimizing the effects of long term trends could be used to choose between these alternative forms of moving average. Limited tests at the close of this study, where Brown's double smoothing model was substituted into Cox's modified model instead of the simple EWMA model normally used, resulted in significant improvements in forecasts of series from the linear generating models. (Excluding the Theil-Wage series, the average specification error was reached from 1.58 to 0.58.)

c. Investigation of "tracking signals." When such misspecification occur such as using the Holt-Winters model

to forecast the Theil-Wage series, a model modification which has been suggested by some [Ref. 6, 16, 17, 18) is the use of a tracking signal to serve as the "exception" to management that the forecast model is not able to perform satisfactorily. The forecast error tolerance of this signal may be set at any level needed. Such signals may be used to obtain the attention of management, or as brought out in the references given, to adjust the smoothing constants to accommodate the series. This procedure probably has the most potential for handling the wide assortment of random process forecasts which is often required. The calculation of "optimal parameters" is obviously not a solution due to the variety of series encountered in practice. Empirical determination of "generally good" parameters such as Winters and Brown have done, fails when gross specification error occurs, but a combination of a "good" parameter and a tracking signal adjustment would appear to be an efficient solution to the problem. As an extension to this thesis it would be interesting to determine the degree of improvement gained by the addition of tracking signals to the models. Although such signals have generally been discussed in the context of EWMA models, the concept very likely could be applied to least squares models and improve their performance in those cases where the model assumptions are violated. Successful application of such refinements would contribute substantially to a more "automatic" forecasting system and increase the population of time series for which any particular model may be successfully applied.

2. Application of Methodology

While it might be premature to suggest that any actions be taken on the sample results obtained in this thesis, it is maintained that the procedure used is sound, and could be applied immediately to a practical problem using actual data. Slight modifications must be made to the Fortran program such as substituting "read" statements for the series generators and a logical check added to determine the model with least forecast error variance for a specific series. The program would then give results which could be readily interpreted by forecast personnel. It is strongly recommended that NavSup or any other agencies using forecast models give consideration to use of this methodology as an evaluation of their existing forecast model. If the quantitative measure obtained justifies continued use of the same model, then the procedure may be repeated using only versions of the same model with varied parameters as a sensitivity evaluation on the model. The small amount of computer time required to perform a comparison (less than 30 seconds on an IBM 360-67) or sensitivity analysis is trivial compared to the potential increase in effectiveness of forecasts if it is discovered that some other model or combination of models are more satisfactory predictors of the random processes of interest.

```
C   A COMPARISON OF SHORT TERM FORECAST MODELS
C
C   THIS PROGRAM GENERATES SEVEN FORMS OF TIME SERIES WHICH
C   ARE EACH FORECAST BY SEVEN FORECAST MODELS, ONE OF WHICH
C   IS THE OPTIMAL(MINIMUM MEAN SQUARED ERROR) PREDICTOR OF
C   THE SERIES.  THE SPECIFICATION ERROR IS COMPUTED FOR
C   EACH FORECAST MODEL, FOR EACH TIME SERIES FORECAST.  THE
C   AVERAGE SPECIFICATION ERROR MEAN AND VARIANCE ARE THEN
C   CALCULATED AS AN OVERALL MEASURE OF FORECAST MODEL PERF.
C
      DIMENSION OBSN(501),RNUM(500),DIFLSE(500),DIFSEA(500),
     1DIFMEA(500),DIFHLT(500),DIFBRN(500),SUMSQ(7),
     2HLTLVL(500),ACOPF(500),DIFF(500),DIFMLS(500),
     3DIFTHL(500),THLLVL(500),SPECF  (7,7),COMBMU(7),
     4SUMER(7),VARSPC(7)
C
C   LOWLUP VALUE IS THE NUMBER OF INITIAL FORECAST PERIODS
C   WHICH ARE DISIGNATED FOR FORECAST STABILIZATION. AFTER
C   THIS PERIOD,FCST ERROR CONTRIBUTES TO VAR. CALCULATION.
C   LOWLUP SHOULD NOT BE LESS THAN 3
      LOWLUP=100
C
C   MAXLUP IS THE TOTAL NUMBER OF OBSERVATIONS GENERATED(OR
C   READ IN IF THE PROGRAM IS MODIFIED TO USE ACTUAL DATA).
C   MAXLUP SHOULD NOT EXCEED 500 WITHOUT PGM MODIFICATION.
      MAXLUP=400
C
C   THIS INITIALIZES THE U(0,1) GENERATOR
      X=URN(-5)
C
C   'NUMBER' SPECIFIES THE TYPE SERIES TO BE GENERATED. 'DO
C   1000' STEPS THROUGH ALL SERIES- 1= LSE SERIES, 2= MOD LSE
C   SERIES, 3= SIMPLE EWMA SERIES, 4= MOD EWMA SERIES (EXPO
C   AUTOCORRELATED), 5= HOLT-WINTERS LINEAR GROWTH SERIES,
C   6= THEIL-WAGE LINEAR GROWTH SERIES, 7= BROWN LINEAR
C   GROWTH SERIES.
      DO 1000 NUMBER=1,7
      OBSN(1)=0.0
      RNUM(1)=0.0
      GO TO (11,66,22,33,44,77,55),NUMBER
C   ****************************************************************
C
C   THIS MODEL GENERATES A SERIES FOR WHICH A ZERO-INTERCEPT
C   LINEAR FORECAST MODEL IS OPTIMAL
C
C   'B' IS THE SLOPE VALUE TO BE SELECTED AS DISIRED.
   11 B=3.0
C
C   RNSIG IS THE STD DEV OF NORMAL RV
      RNGSIG=3.0
C   THIS DO LOOP GENERATES THE DESIRED NUMBER OF RNS
      DO 100 I=1,MAXLUP
      SUM=0.0
      DO 20 J=1,12
C   THIS IS THE UNIF(0,1) RN GENERATOR
      X=URN(2)
      R=X
   20 SUM=SUM+R
C
C   GENERATOR WHICH PRODUCES A NORMAL(0,X) RANDOM NUMBER.
      RNUM(I)=(SUM-6.0)*RNGSIG
C   THIS COMPUTES THE DETERMINISTIC PART OF THE LINEAR MODEL
C   AND ADDS THE NORMAL RV TO IT.
      OBSN(I)=B*I+RNUM(I)
  100 CONTINUE
      GO TO 60
C   ****************************************************************
C
C   THIS MODEL TAKES SAME NORMAL RN'S USED ABOVE AND GENERATE
C   A FIRST ORDER AUTOREGRESSIVE SERIES FOR WHICH THE MOD
C   LEAST SQUARES FORICAST IS OPTIMAL.
C
```

```
   66 RHO=0.4
      EPS=0.0
      DO 101 I=1,MAXLUP
C  THIS GENERATES CORRELATED RANDOM SHOCKS. SEE EQ.(A-4)
      EPS=RHO*EPS+RNUM(I)
C  THIS GENERATES THE DETERMINISTIC PART OF THE LINEAR MODEL
C  AS BEFORE, BUT NOW THE CORRELATED SHOCK IS ADDED.
      OBSN(I)=B*I+EPS
      RNUM(I)=EPS
  101 CONTINUE
      GO TO 60
C  ***********************************************************
C
C  THIS MODEL GENERATES A SERIES FOR WHICH THE SIMPLE EWMA
C  FORECAST MODEL IS OPTIMAL.
C
C  'RNGGAM' IS THE RANGE OF RANDOMNESS ASSOCIATED
C  WITH THE SHOCK TO PROCESS LEVEL  SHOULD BE SET LESS THAN
C  THE DEGREE ,RNG,ASSOCIATED WITH  THE RANDOM SHOCK ADDED.
   22 RNGGAM=5.0
C
C  DEGREE OF RANDOMNESS,RNG ,OF UNIF RV SHOULD ALSO BE SET
      RNG=12.0
      MU=0.0
C  THIS IS A UNIF(-X.0,X.0) RNDM GENERATOR. LOOP PROVIDES
C  THE NUMBER OF RNS DESIRED.
      DO 300 I=1,MAXLUP
      X=URN(2)
      RNUM(I)=(X-0.5)*RNG
C  AN IMA PROCESS IS GENERATED USING THE RANDOM SHOCK FORM
C  OF THE MODEL SHOWN IN EQ.(B-4).
      OBSN(I)=MU+RNUM(I)
      X=URN(2)
      BIT=(X-0.5)*RNGGAM
      MU=MU+BIT
      RNUM(I)=RNUM(I)+BIT
  300 CONTINUE
      GO TO 60
C  ***********************************************************
C
C  THIS MODEL GENERATES A SERIES FOR WHICH COX'S MODIFIED
C  FORECAST MODEL IS OPTIMAL. COR COEFF=RHO**LAG. HERE LAG=1
C
C  'RHO1' IS THE DESIRED CORRELATION COEFFICIENT
   33 RHO1=0.4
      CRHO=1.0-RHO1
      UNIRNG=10.0
C  THIS IS A UNIF(-X,.X) RNDMN GENERATOR. LOOP GENERATES
C  THE PROPER NUMBER OF RNS DESIRED.
      DO 200 I=1,MAXLUP
      X=URN(2)
      RNUM(I)=(X-0.5)*UNIRNG
      OBSN(I+1)=RHO1*OBSN(I)+CRHO*RNUM(I)
  200 CONTINUE
      GO TO 60
C  ***********************************************************
C
C  THIS MODEL GENERATES A SERIES FOR WHICH THE HOLT TWO-PARA
C  METER MODEL IS OPTIMAL. SLOPE EXPERIENCES A RNDM CHANGE.
C  THERE IS ALSO A RNDM MOVEMENT ABOUT THE LEVEL OF THE
C  PROCESS CONTRIBUTING TO THE FORECAST ERROR
C
C  THE DEGREE OF RANDOMNESS MUST BE SET FOR THE
C  RANGE OF THE FORECAST ERROR, RNGERR
   44 RNGERR=10.0
      SLOPE=1.0
      VECMU=0.0
      PCSLVL=0.0
      ALFA=0.2
      ALFA1=0.1
      DO 400 I=2,MAXLUP
      X=URN(2)
```

```
C   THIS GENERATES THE U(-X,X) FCST ERROR
        RNUM(I)=(X-0.5)*RNGERR
C   THIS COMPUTES THE OBSERVED VALUE OF THE PROCESS WITH A
C   RANDOM FORECAST ERROR INCLUDED.
        OBSN(I)=PCSLVL+RNUM(I)
        VECMU=VECMU+SLOPE+ALFA*RNUM(I)
C   THIS COMPUTES NEW VALUE OF SLOPE DUE TO RNDM PROCESS CHNG
        SLOPE=SLOPE+ALFA*ALFA1*RNUM(I)
C   THIS COMPUTES NEW LEVEL FOR SERIES BASED ON LATEST EST OF
C   SLOPE PLUS THE RNDM CHANGE FOR THE PERIOD
        PCSLVL=VECMU+SLOPE
  400 CONTINUE
        GO TO 60
C   ******************************************************************
C
C   THIS MODEL GENERATES A SERIES FOR WHICH THE THEIL-WAGE
C   FORECAST MODEL IS OPTIMAL.
C
C   THE RANGE OF UNCERTAINTY OF  FCST ERROR AND SLOPE  CHNGE
C   MUST BE SET TO PROPERLY SPECIFY THE UNDERLYING PROCESS.
   77 YLVL=0.0
        SLP=0.0
        EPSRNG=10.0
        DELRNG=4.0
C   THE GENERATING MODEL CORRESPONDS TO EQ. (D-1)
        DO 450 I=1,MAXLUP
        X=URN(2)
        BIT1=(X-0.5)*DELRNG
        SLP=SLP+BIT1
        YLVL=YLVL+SLP
        X=URN(2)
        BIT2=(X-0.5)*EPSRNG
        OBSN(I)=YLVL+BIT2
        RNUM(I)=BIT1+BIT2
  450 CONTINUE
        GO TO 60
C   ******************************************************************
C
C   THIS MODEL GENERATES A SERIES FOR WHICH BROWN'S ONE
C   PARAMETER MODEL IS OPTIMAL
C
C   THE FOLLOWING INITIAL CONDITIONS AND PARAMS MUST BE SET.
   55 RNGERR=10.0
        ALFA=0.1
        BSLOPE=1.0
        BRNLVL=0.0
        FCSBRN=0.0
        BETA=1.0-ALFA
        BETASQ=BETA*BETA
        ALFASQ=ALFA*ALFA
        DO 500 I=1,MAXLUP
        X=URN(2)
        RNUM(I)=(X-0.5)*RNGERR
        OBSN(I)=FCSBRN+RNUM(I)
C   THIS CONTRIBUTES A RANDOM CHANGE TO PROCESS LEVEL
        BRNLVL=BRNLVL+BSLOPE+(1.0-BETASQ)*RNUM(I)
C   THIS CONTRIBUTES A RANDOM CHANGE TO SLOPE
        BSLOPE=BSLOPE+ALFASQ*RNUM(I)
C   THIS LEVEL CONTRIBUTES TO NEXT PERIOD OBSERVATION
        FCSBRN=BRNLVL+BSLOPE
  500 CONTINUE
C   ******************************************************************
C
C   THIS  SECTION  COMPUTES MEAN, VARIANCE & AUTOCORRELATION
C   OF THE OBSERVATIONS. THE CORRELATION BETWEEN SUCCESSIVE
C   OBSERVATIONS, RHO1, IS USED IN SUBSEQUENT FORECAST MODELS
C
   60 NOBSN=MAXLUP
        MAXLAG=5
        K=2
    3 GO TO (5,4,99),K
    5 K=3
```

```
      WRITE(6,2499)NUMBER
 2499 FORMAT(1H ,/////,10X,'STATISTICAL CHARACTERISTICS OF',
     1' THE GENERATED SERIES',5X,'NUMBER= ',I3,//)
      GO TO 9
    4 K=1
      WRITE(6,2498)NUMBER
 2498 FORMAT(1H ,/////,10X,'STATISTICAL CHARACTERISTICS OF',
     1' THE RNDM SERIES',5X,'NUMBER= ',I3,//)
      XSUM=0.0
      DO 12 I=1,MAXLUP
   12 XSUM=XSUM+RNUM(I)
      XBAR=XSUM/(NOBSN*1.0)
      VAR=0.0
      DO 21 I=1,NOBSN
      DIFF(I)=RNUM(I)-XBAR
   21 VAR=DIFF(I)**2+VAR
      GO TO 24
    9 XSUM=0.0
      DO 10 I=1,MAXLUP
   10 XSUM=OBSN(I)+XSUM
      XBAR=XSUM/(NOBSN*1.0)
      VAR=0.0
      DO 23 I=1,NOBSN
      DIFF(I)=OBSN(I)-XBAR
   23 VAR=DIFF(I)*DIFF(I)+VAR
   24 VAR=VAR/(NOBSN*1.0)
C  VAR IS VARIANCE OR AUTOCOVARIANCE FCTN WITH ZERO LAG.
   25 DO 40 J=1,MAXLAG
C  LIMLUP IS THE UPPER LOOP LIMIT WHICH DECREASES WITH LAG.
      LIMLUP=NOBSN-J
      ACVF=0.0
      DO 30 I=1,LIMLUP
      XPROD=DIFF(I)*DIFF(I+J)
   30 ACVF   =XPROD+ACVF
      ACVF   =ACVF   /(NOBSN*1.0)
C  ACVF IS THE AUTOCOVARIANCE FUNCTION WITH LAG J.
      ACORF(J)=ACVF   /VAR
C  ACORF IS THE AUTOCORRELATION FCTN, THE NORMALIZED ACVF.
      WRITE(6,2501)J,ACORF(J)
 2501 FORMAT(1H,/10X,'AUTOCORRELATION COEFFICIENT WITH LAG'
     1,I3,'= ',F7.5)
   40 CONTINUE
      RHO1=ACORF(1)
      WRITE(6,2600)XBAR,VAR,RHO1
 2600 FORMAT(1H,/5X,'XBAR=',F10.5,5X,'VAR=',F10.5,5X,'RHO1='
     1,F7.5,//)
      GO TO 3
C  ***********************************************************
C
C  THIS IS THE LSE FORECAST MODEL (ZERO INTERCEPT FORM)
C
C  THESE INITIAL CONDITIONS SPECIFICALLY ANTICIPATE THE
C  LSE SERIES GENERATED EARLIER.
   99 AHAT=0.0
      SUMXY=B
      SUMXSQ=1.0
      DIFLSE(1)=0.0
      FCSLSE=B*2.0
      FCSMLS=0.0
      RHO=0.4
      DO 600 I=2,MAXLUP
C  FORECAST ERROR IS CALCULATED AND ACCUMULATED FOR LATER.
      DIFLSE(I)=FCSLSE   -OBSN(I)
      DIFMLS(I)=FCSMLS-OBSN(I)
C  BEFORE COEFFICIENTS CAN BE ESTIMATED,XBAR AND YBAR MUST
C  BE CALCULATED.
      SUMXY=SUMXY+I*OBSN(I)
      SUMXSQ=SUMXSQ+I*I
      BHAT=SUMXY/SUMXSQ
C  NOW THE FCST CAN BE MADE,USING THE COEFFICIENTS,AHAT&BHAT
      FCSLSE=AHAT+BHAT*(I+1)
C
```

```
C
C   THIS IS THE MODIFIED LEAST SQUARES FORECAST MODEL.
C   SINCE IT USES THE SAME PARAMETER ESTIMATES AS THE LSE
C   MODEL PLUS A CORRECTION, ITS FORECASTS ARE GENERATED AT
C   THE SAME TIME AS THE LSE MODEL. THIS IS THE SAME AS EQ.
C   (A-6),EXCEPT SIGN CHANGE DUE TO DIFFERENT FCST ERROR FORM
        FCSMLS=FCSLSE-RHO*DIFLSE(I)
  600 CONTINUE
C   ***********************************************************
C
C   THIS IS THE SIMPLE EWMA FORECAST MODEL WITH OPTIMAL ALFA
C
C   FCS1 IS THE INITIAL FORECAST NEEDED TO START THE SIMPLE
C   EXPONENTIALLY WEIGHTED MOVING AVERAGE FORECAST SERIES
        FCS1=0.0
        ALFA=(3.0*RHO1-1.0)/(2.0*RHO1)
        IF(RHO1.LE.0.3333) ALFA=0.075
        BETA=1.0-ALFA
        DO 700 I=1,MAXLUP
        DIFSEA(I)=FCS1-OBSN(I)
C   THIS GENERATES FORECAST FOR NEXT PERIOD USING SIMPLE EWMA
        FCS1=ALFA*OBSN(I)+BETA*FCS1
  700 CONTINUE
C   ***********************************************************
C
C   THIS IS THE MODIFIED EWMA FORECAST MODEL PROPOSED BY COX
C
        ALFA=0.01
        BETA=1.0-ALFA
        FCSEMA=0.0
        DO 800 I=2,MAXLUP
C   THIS GENERATES FORECAST FOR NEXT PERIOD USING SIMPLE EWMA
C   AND OPTIMAL ALFA ASSUMING EXPONENTIAL AUTOCORRELATION.
        FCSEMA    =ALFA*OBSN(I-1)+BETA*FCSEMA
C   THIS USES SIMPLE EWMA IN COX'S MODIFIED EWMA.
        FCSMEA    =RHO1*OBSN(I-1)+(1.0-RHO1)*FCSEMA
        DIFMEA(I)=FCSMEA    -OBSN(I)
  800 CONTINUE
C   ***********************************************************
C
C   THIS IS THE HOLT TWO-PARAMETER FORECAST MODEL
C
        ALFA=0.2
        BETA=1.0-ALFA
        BSLOPE=0.0
        HLTLVL(1)=0.0
        FCSHLT=0.0
        ALFA1=0.1
        BETA1=1.0-ALFA1
        DO 900 I=2,MAXLUP
        DIFHLT(I)=FCSHLT    -OBSN(I)
C   THIS ESTIMATES THE CURRENT PROCESS LEVEL
        HLTLVL(I)=ALFA*OBSN(I)+BETA*(HLTLVL(I-1)+BSLOPE)
C   SLOPE IS UPDATED FOR USE ON NEXT ITERATION.
        BSLOPE    =ALFA1*(HLTLVL(I)-HLTLVL(I-1))+BETA1*BSLOPE
C   FORECAST IS GENERATED USING CURRENT LEVEL  AND  SLOPE EST
        FCSHLT=HLTLVL(I)+BSLOPE
  900 CONTINUE
C   ***********************************************************
C
C   THIS IS THE THEIL-WAGE  FORECAST MODEL.
C
        FCSTHL=0.0
        ALFA=0.596
        ALFAP=0.425
        BETA=1.0-ALFA
        BETAP=1.0-ALFAP
        FACALF=ALFA/(1.0-BETA*ALFAP)
        FACBET=BETA*BETAP/(1.0-BETA*ALFAP)
        THLLVL(1)=0.0
        SLP=0.0
        DO 925 I=2,MAXLUP
```

```
      DIFTHL(I)=FCSTHL-OBSN(I)
      THLLVL(I)=FACALF*OBSN(I)+FACBET*(THLLVL(I-1)+SLP)
      SLP=ALFAP*(THLLVL(I)-THLLVL(I-1))+BETAP*SLP
      FCSTHL=THLLVL(I)+SLP
  925 CONTINUE
C   **************************************************************
C
C   THIS IS BROWN'S ONE-PARAMETER FORECAST MODEL
C
      ALFA=0.1
      BETA=1.0-ALFA
      SNGLSM=0.0
      DBLSMO=0.0
      BSLOPE=0.0
      FCSBRN=0.0
      DIFBRN(1)=0.0
      FACTOR=ALFA/BETA
      DO 950 I=2,MAXLUP
      DIFBRN(I)=FCSBRN-OBSN(I)
C   THIS COMPUTES THE SINGLE (1ST ORDER) SMOOTHED ESTIMATE
      SNGLSM    =ALFA*OBSN(I)+BETA*SNGLSM
C   THIS COMPUTES THE 2ND ORDER SMOOTHED ESTIMATE OF SERIES
      DBLSMO    =ALFA*SNGLSM    +BETA*DBLSMO
      SLOPE    =FACTOR*(SNGLSM    -DBLSMO    )
      FCSBRN        =2.0*SNGLSM    -DBLSMO+ SLOPE
  950 CONTINUE
C   **************************************************************
C
      DO 2999 J=1,7
      SUMER(J)=0.0
 2999 SUMSQ(J)=0.0
C   THIS COMPUTES SUMS OF FORECAST ERRORS AND SQUARED ERRORS.
      DO 3000 I=LOWLUP,MAXLUP
      SUMER(1)=SUMER(1)+DIFLSE(I)
      SUMSQ(1)=SUMSQ(1)+DIFLSE(I)**2
      SUMER(2)=SUMER(2)+DIFMLS(I)
      SUMSQ(2)=SUMSQ(2)+DIFMLS(I)**2
      SUMER(3)=SUMER(3)+DIFSEA(I)
      SUMSQ(3)=SUMSQ(3)+DIFSEA(I)**2
      SUMER(4)=SUMER(4)+DIFMEA(I)
      SUMSQ(4)=SUMSQ(4)+DIFMEA(I)**2
      SUMER(5)=SUMER(5)+DIFHLT(I)
      SUMSQ(5)=SUMSQ(5)+DIFHLT(I)**2
      SUMER(6)=SUMER(6)+DIFTHL(I)
      SUMSQ(6)=SUMSQ(6)+DIFTHL(I)**2
      SUMER(7)=SUMER(7)+DIFBRN(I)
 3000 SUMSQ(7)=SUMSQ(7)+DIFBRN(I)**2
C
C   THIS COMPUTES ESTIMATED FORECAST ERROR MEAN AND VARIANCE
C   FOR EACH FORECAST MODEL.
      VARFAC=MAXLUP-LOWLUP-1.0
      DO 4000 J=1,7
      SUMER(J)=SUMER(J)/(VARFAC+1.0)
 4000 SUMSQ(J)=SUMSQ(J)/VARFAC
C
C   THIS COMPUTES THE MEASUREMENT OF SPECIFICATION ERROR,
C   THE RATIO OF FORECAST MODEL'S FORECAST ERROR VARIANCE
C   TO THE VARIANCE OF THE OPTIMAL MODEL FOR THAT SERIES.
      DO 5000 J=1,7
 5000 SPECF    (NUMBER,J)=(SUMSQ(J)/SUMSQ(NUMBER))-1.0
      WRITE(6,5001)(L,SPECF    (NUMBER,L),L,SUMSQ(L),L,SUMER
     1(L),L=1,7)
 5001 FORMAT(1H ,/,5X,'SPEC ERROR (',I2,' )=',2X,F10.5,
     1 5X,'VAR(',I2,' )=',2X,F12.4,5X,'AVG ERR(',I2,' )=',
     22X,F10.5,/)
C
C   THIS PRINTS OUT THE OBSERVATIONS AND THE FORECAST ERRORS
C   EXPERIENCED BY EACH FORECAST MODEL.
      WRITE(6,2099)NUMBER
 2099 FORMAT(1H ,////,7X,'SUMMARY OF FORECAST RESULTS FOR ',
     1'THE GENERATED SERIES',5X,'NUMBER= ',I3,///)
      WRITE(6,2100)
```

```
 2100 FORMAT(1H ,//,5X,'SERIES OBSN',4X,'DIFF L SQ. ',4X,
     1'DIFF M LSE',5X,'DIFF SEWMA',5X,'DIFF MEWMA',5X,
     2'DIFF HOLT ',5X,'DIFF THIEL',5X,'DIFF BROWN', /)
      WRITE(6,2200)(OBSN(L),DIFLSE(L),DIFMLS(L),DIFSEA(L),
     1DIFMEA(L),DIFHLT(L),DIFTHL(L),DIFBRN(L),L=LOWLUP,
     2MAXLUP)
 2200 FORMAT(1H ,8(5X,F10.4))
 1000 CONTINUE
      NUMBER=7
      FNUM=NUMBER*1.0
C
C  THIS COMPUTES THE AVERAGE SPEC ERROR FOR EACH OF THE
C  FORECAST MODELS,J,OVER ALL SERIES,I.
      DO 1100 J=1,NUMBER
      COMSUM=0.0
      DO 1111 I=1,NUMBER
 1111 COMSUM=COMSUM+SPECE(I,J)
      COMBMU(J)=COMSUM/FNUM
C
C  THIS COMPUTES THE SPECIFICATION ERROR SAMPLE VARIANCE
C  FOR EACH FORECAST MODEL,J,OVER ALL SERIES,I.
      ESQSUM=0.0
      DO 1112 I=1,NUMBER
 1112 ESQSUM=ESQSUM+(SPECE(I,J)-COMBMU(J))**2
      VARSPC(J)=ESQSUM/FNUM
 1100 CONTINUE
C
C  THIS SUMMARIZES THE SPECIFICATION ERROR DATA FOR COMPAR.
      WRITE(6,3100)
 3100 FORMAT(1H ,////,20X,'LSE MDL',8X,'MLSE MDL' 7X,'EWMA M
     17X,'MEWMA MDL',6X,'HLT-WNT MDL',4X,'THL-WGE MDL',4X,
     2'BRN MDL'/)
      DO 3111 I=1,NUMBER
 3111 WRITE(6,3110) I,(SPECE(I,J),J=1,NUMBER)
 3110 FORMAT(1H ,  6X,'SERIES(',I2,')',2X,7(F9.5,6X) //)
      WRITE(6,3050)
 3050 FORMAT(1H ,////,10X,'THIS SUMMARIZES SPEC ERROR FOR ',
     1'COMPARISON',/)
      WRITE(6,3120) (L,COMBMU(L),VARSPC(L),L=1,NUMBER)
 3120 FORMAT(1H ,10X,'FORECAST MODEL (',I2,')',5X,F9.5,10X,
     1F10.5)
      STOP
      END
```

LIST OF REFERENCES

1. Zehna, P. W., Probability Distributions and Statistics, p. 477-501, Allyn and Bacon, Inc., 1970.

2. Coventry, J. A., A Comparison of Demand Forecasting Techniques, M.S. Thesis, US Naval Postgraduate School, March 1971.

3. Box, G. E. P. and Jenkins, B. M., Time Series Analysis, Forecasting and Control, p. 103-108, Holden-Day, Inc., 1970.

4. Muth, J. F., "Optimal Properties of Exponentially Weighted Forecasts," Journal of the American Statistical Association, p. 299-306, June 1960.

5. Cox, D. R., "Prediction By Exponentially Weighted Moving Averages and Related Methods," Journal of the Royal Statistical Society, B23, p. 414-442, 1966.

6. Brown, R. G., Smoothing, Forecasting and Prediction, Prentice Hall Inc., 1963.

7. Harrison, P. J., "Exponential Smoothing and Short term Sales Forecasting," Management Science, v. 13, No. 11, July 1967.

8. Naylor, T. H., and others, Computer Simulation Techniques, p. 118-121, Wiley, 1968.

9. Holt, C. C., Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.

10. Winters, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages," Management Science, v. 6, No. 3, p. 324-342, April 1960.

11. Theil, H. and Wage, S., "Some Observations on Adaptive Forecasting," Management Science, v. 10, No. 2, January 1964.

12. Johnston, J., Econometric Methods, p. 177-199, McGraw-Hill Inc., 1963.

13. Bossons, J., "The Effects of Parameter Misspecification and Nonstationarity on the Applicability of Adaptive Forecasts," Management Science, vol. 12, No. 9, May 1966.

14. Box, G. E. P. and Jenkins, G. M., "Some Statistical Aspects of Adaptive Optimization and Control," Journal of the Royal Statistical Society, B, 24 (2), p. 297, 1962.

15. Gilchrist, W. G., "Methods of Estimation Involving Discounting," Journal of the Royal Statistical Society, B, 29, No. 2, p. 355-369, 1967.

16. Trigg, D. W., "Monitoring a Forecasting System," Operational Research Quarterly, v. 15, p. 271-274, 1964.

17. Trigg, D. W. and Leach, A. G., "Exponential Smoothing with an Adaptive Response Rate," Operational Research Quarterly, v. 18, No. 1, p. 53-59, 1967.

18. Rao, A. G. and Shapiro, A., "Adaptive Smoothing Using Evolutionary Spectra," Management Science, v. 17, No. 3, p. 208-218, November 1970.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Chief of Naval Personnel Pers 11b Department of the Navy Washington, D. C. 20370	1
4.	Naval Postgraduate School Department of Operations Research and Administrative Sciences Monterey, California 93940	1
5.	Professor P. W. Zehna, Code 55Ze Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
6.	MAJ Ralph E. Hayes, USA 3902 Commander Drive Columbus, Georgia 31903	1

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Naval Postgraduate School
Monterey, California

2a. REPORT SECURITY CLASSIFICATION

UNCLASSIFIED

2b. GROUP

3. REPORT TITLE

A COMPARISON OF SHORT-TERM FORECASTING MODELS

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; September 1971

5. AUTHOR(S) (First name, middle initial, last name)

Ralph E. Hayes

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1971	95	18

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

Seven short-term forecasting models, two using least-squares estimation methods and five employing variations of the exponentially weighted moving average method, are compared in their relative ability to produce minimum error variance forecasts for seven simulated time series. Each series was generated to enable one of the forecast models to be the least squared error predictor. A comarison methodology is developed which facilitates forecast model performance through the measurement of model specification errors. A computer program is presented which may be modified to accept real time series and which permits the forecast models to be ranked in order of their relative specification error.

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Forecasting Models						
Model Specification Error Measurement						
Time Series Simulation						
Forecast Model Performance Comparison						
Adaptive Forecasting						
Prediction Methods						
Exponential Smoothing						
Estimation						

Thesis 131452
H4056 Hayes
c.1 A comparison of
short-term forecasting
models.